BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Appendix I
(Copies of the following references are submitted along with the instant response)

1. Widespread occurrence of three sequence motifs in diverse S-Adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Archives of Biochemistry and Biophysics* 310 (2), 417-427, 1994 (common structure was found in 69 methyltransferases).

2. Crystal structure of the chemotaxis receptor methyltransferase CheR suggests a conserved structural motif for binding S-Adenosylmethionine. *Structure* 5, 545-558, 1997 (protein methyltransferase was found to share the same catalytic domain with DNA, RNA and small molecule methyltransferases).

3. Structure of PvuII DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Research* 25 (14), 2702-2715, 1997 (DNA- methyltransferase *Pvu*ll had identical binding domains for SAM as did M-Taq1 and M-Hha1).

ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 316, No. 1, January 10, p. 657, 1995

ERRATUM

Volume **310,** Number 2 (1994), in the article "Widespread Occurrence of Three Sequence Motifs in Diverse *S*-Adenosylmethionine-Dependent Methyltransferases Suggests a Common Structure for These Enzymes," by Ron M. Kagan and Steven Clarke, pages 417–427: On page 425, the sequence in Table III identified as *E. coli* AdoMet decarboxylase (and discussed on page 422 of the text) is in fact that of spermidine synthase, encoded by the *E. coli speE* gene (GenBank Accession No. J02804). Sequence analysis of *E. coli* AdoMet decarboxylase does not demonstrate any of the three methyltransferase motifs as stated and thus lessens the chances that the motif III-like sequences of the human, rat, and hamster AdoMet decarboxylases described in Table III are significant. Therefore, it does not appear that any AdoMet decarboxylase shares sequence motifs with the group of methyltransferases described in the paper. On the other hand, it is now clear that spermidine synthase, from both *E. coli* and human (Genbank Accession No. M64231) sources, does have the three sequence motifs characteristic of many *S*-adenosylmethionine-dependent methyltransferases. This enzyme catalyzes propylamine transfer from decarboxylated *S*-adenosylmethionine to putrescine. Interestingly, both spermidine synthases share sequence similarity to a recently identified *S*-adenosylmethionine-dependent putrescine *N*-methyltransferase (Genbank Accession No. D28506). The authors thank Professor Anthony Pegg (Pennsylvania State University) for helping bring these facts to their attention and regret any confusion caused by their error.

Archives of
Biochemistry and Biophysics

Volume 310, Number 2, May 1, 1994

CONTENTS BY SHORT TITLES

RESEARCH REPORT

CD44 K-ras, and p53 in Colorectal Neoplasia 504

REGULAR ARTICLES

Nucleotide Recognition by Histone H1 291
Cleavage of Epitectin by Bovine *Pasteurella haemolytica* Glycoprotease 300
Conformational Studies of a Nebulin Module 310
Electron Transfer from Cytochrome P450 Reductase to Cytochrome b_5 318
Rat Liver Aryl Sulfotransferase Sulfation of Polyaromatic Phenols 325
Insect P450 Isozymes Selectively Metabolize Linear Furanocoumarins 332
Tris Hydroxamate Mn(III) Catalysts of O_2^- Dismutation 341
1,25-Dihydroxyvitamin D_3 Receptor 347
Endothelial Cell Peroxynitrite Production 352
Acetobacter hansenii NAD-Preferring Glucose 6-Phosphate Dehydrogenase 360
Cytochrome P450-Cytochrome b_5 Interactions 367
Lacto-Series Gangliosides from Bovine Erythrocytes 373
Protein Binding of Muconaldehyde 385
Protein Electron Transfer at a Membrane Surface 392
Roles of Glu318 and Thr319 in CYP1A2 Catalysis 397
Novel Metabolic Pathway of Arylethers by Cytochrome P450 402
Coexistence of Discrete Forms of Mitochondrial Hexokinase 410
Methyltransferase Sequence Motifs 417
Binding of Heat-Shock Protein 70 to Tubulin 428
Disulfide Bonds in PDGF-AA 433
Heme Protein Dynamics 440
Oxidation of Dimethylselenide by Flavin-Containing Monooxygenase 448
NADPH-Cytochrome P450 Oxidoreductase Promoter 452
High-Resolution Structure of the *Paracoccus denitrificans* Cytochrome c_2 460
Bovine Heart Fructose 6-*P*,2-kinase:Fructose 2,6-Bisphosphatase Gene 467
Lys^{331} and Gly^{15} Mutants of Adenylosuccinate Synthetase 475
Regulation of Human Monocyte Metalloproteinases by G-Proteins 481
Low-Density Lipoprotein Aggregation by Azo Initiator 489
Chicken Liver Alanine:2-Oxoglutarate (Glyoxylate) Aminotransferase 497

NOTICE

The Subject Index for Volume 310 will appear in the December 1994 issue as part of a cumulative index for the year 1994.

Archives of
Biochemistry and Biophysics

Volume 310, Number 2, May 1, 1994

0003-9861/94 $6.00

MADE IN THE UNITED STATES OF AMERICA

This journal is printed on acid-free paper.

ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS

(ISSN 0003-9861)

Published monthly (except semimonthly in February, May, August, and November) by Academic Press, Inc., 6277 Sea Harbor Drive, Orlando, FL 32887-4900

1994: Volumes 308–315. Price: $1392.00 U.S.A. and Canada; $1645.00 all other countries
All prices include postage and handling.

All correspondence and subscription orders should be addressed to the office of the Publishers at 6277 Sea Harbor Drive, Orlando, FL 32887-4900. Send notices of change of address to the office of the Publishers at least 6 to 8 weeks in advance. Please include both old and new addresses. POSTMASTER: Send changes of address to *Archives of Biochemistry and Biophysics*, 6277 Sea Harbor Drive, Orlando, FL 32887-4900.

Second class postage paid at Orlando, FL, and at additional mailing offices.

ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 310, No. 2, May 1, pp. 417–427, 1994

Widespread Occurrence of Three Sequence Motifs in Diverse S-Adenosylmethionine-Dependent Methyltransferases Suggests a Common Structure for These Enzymes

Ron M. Kagan and Steven Clarke[1]

Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California, Los Angeles, California 90024-1569

Received October 19, 1993, and in revised form December 21, 1993

Three regions of sequence similarity have been reported in several protein and small-molecule S-adenosylmethionine-dependent methyltransferases. Using multiple alignments, we have now identified these three regions in a much broader group of methyltransferases and have used these data to define a consensus for each region. Of the 84 non-DNA methyltransferase sequences in the GenBank, NBRF PIR, and Swissprot databases comprising 37 distinct enzymes, we have found 69 sequences possessing motif I. This motif is similar to a conserved region previously described in DNA adenine and cytosine methyltransferases. Motif II is found in 46 sequences, while motif III is found in 61 sequences. All three regions are found in 45 of these enzymes, and an additional 15 have motifs I and III. The motifs are always found in the same order on the polypeptide chain and are separated by comparable intervals. We suggest that these conserved regions contribute to the binding of the substrate S-adenosylmethionine and/or the product S-adenosylhomocysteine. These motifs can also be identified in certain nonmethyltransferases that utilize either S-adenosylmethionine or S-adenosylhomocysteine, including S-adenosylmethionine decarboxylase, S-adenosylmethionine synthetase, and S-adenosylhomocysteine hydrolase. In the latter two types of enzymes, motif I is similar to the conserved nucleotide binding motif of protein kinases and other nucleotide binding proteins. These motifs may be of use in predicting methyltransferases and related enzymes from the open reading frames generated by genomic sequencing projects. © 1994 Academic Press, Inc.

Of the 3196 enzymes described in the latest version of *Enzyme Nomenclature* (1), about 3% represent species that catalyze the attack of a variety of nitrogen, oxygen, carbon, and sulfur nucleophiles on the methyl group of S-adenosylmethionine. These methyltransferases include enzymes that result in the formation of methyl ester, methyl ether, methyl thioether, methyl amine, methyl amide and other derivatives on proteins, nucleic acids, polysaccharides, lipids, and various small molecules. We have been interested in the possibility that the similarity in the simple catalytic chemistry of these reactions is reflected in amino acid sequence and three-dimensional structural similarities of the enzymes. Since all of these enzymes bind the methyl donor AdoMet[2] and produce AdoHcy as a product, a similar binding pocket in these enzymes may be reflected in similar amino acid sequences. Such a situation is found, for example, in the GTP-binding proteins which possess three short sequence motifs with distinct spacing that interact with GTP (2, 3).

The first indications of sequence similarities between different types of methyltransferases were shown in enzymes that modify DNA (4, 5). For example, DNA m^5C methyltransferases have 10 regions of sequence similarity, two of which are also shared with DNA m^6A methyltransferases (5, 6). It was later shown that at least one of these motifs (termed motif I in this work) is also shared with several RNA, protein, and small molecule methyltransferases and with AdoHcy hydrolases (7). Two additional sequence motifs (termed motif II and motif III in this work) are found in a number of protein and small molecule methyltransferases, in at least one tRNA methyltrans-

[1] To whom correspondence should be addressed.

[2] Abbreviations: AdoMet, S-adenosyl-L-methionine; AdoHcy, S-adenosyl-L-homocysteine.

0003-9861/94 $6.00

TABLE I

Methyltransferase Sequences Examined in This Study

	Enzyme	Gene[a]	EC Number	Organism	Motifs[b]	Accession[c]
Protein carboxyl MTases	Isoaspartyl O-MT	PIMT	2.1.1.77	human	I,II,III	*A33404*
	Isoaspartyl O-MT	PIMT	2.1.1.77	bovine	I,II,III	A34242
	Isoaspartyl O-MT	PIMT	2.1.1.77	mouse	I,II,III	M60320
	Isoaspartyl O-MT	PIMT	2.1.1.77	rat	I,II,III	D11475
	Isoaspartyl O-MT	PIMT	2.1.1.77	wheat	I,II,III	L07941
	Isoaspartyl O-MT	*pcm*	2.1.1.77	*E. coli*	I,II,III	M63493
	γ-Glutamyl O-MT	*cheR*	2.1.1.80	*E. coli*	I,II,III	M13463
	γ-Glutamyl O-MT	*cheR*	2.1.1.80	*S. typhimurium*	I,II,III	J02757
	γ-Glutamyl O-MT	*frzF*	2.1.1.80	*M. xanthus*	I,II,III	M35200
	Isoprenylcysteine O-MT	*STE14*	2.1.1.100	yeast	nd	L15442
Small molecule O-MTases	Acetylserotonin O-MT	HOMT	2.1.1.4	human	I,II,III	M83779
	Acetylserotonin O-MT	HOMT	2.1.1.4	bovine	I,II,III	J02671
	Acetylserotonin O-MT	HOMT	2.1.1.4	chicken	I,II,III	X62309
	Catechol O-MT	COMT	2.1.1.6	human	I,II,III	M65212
	Catechol O-MT	COMT	2.1.1.6	rat	I,II,III	M60754
	Caffeic acid O-MT	(CAOMT)	2.1.1.68	maize	I,II,III	M73235
	Caffeic acid O-MT	(CAOMT)	2.1.1.68	alfalfa	I,II,III	M63853
	Caffeic acid O-MT	(CAOMT)	2.1.1.68	aspen	I,II,III	X62096
	Caffeoyl CoA O-MT	(CCOMT)	2.1.1.104	parsley	I,II,III	M69184
	O-Demethyl puromycin O-MT	*dmpM*	2.1.1.38	*S. alboniger*	I,II,III	M74560
	Hydroxyneurosporene O-MT	*crtF*		*R. capsulatus*	I,II,III	*S04408*
	Myo-inositol O-MT	*Imt1*	2.1.1.39/40	*M. crystallinum*	I,II,III	M87340
	Carminomycin O-MT	*dnrK*		*S. peucetius*	I,II,III	L13453
	Tetracenomycin 3-O-MT	*tcmN*		*S. glaucesens*	I,II,III	M80674
	Tetracenomycin 8-O-MT	*tcmO*		*S. glaucesens*	I,II,III	M80674
	Midamycin O-MT	*mdmC*		*S. mycarofaciens*	I,II,III	M93958
	Erythromycin biosynthesis O-MT	*eryG*		*S. erythraea*	I,II,III	*S18533*
Small molecule N-MTases	Phenylethanolamine N-MT	PNMT	2.1.1.28	human	I,II,III	J03280
	Phenylethanolamine N-MT	PNMT	2.1.1.28	bovine	I,II,III	M36706
	Phenylethanolamine N-MT	PNMT	2.1.1.28	rat	I,II,III	X14211
	Phenylethanolamine N-MT	PNMT	2.1.1.28	mouse	I,II,III	L12687
	Glycine N-MT	GNMT	2.1.1.20	rabbit	I,II,III	D13307
	Glycine N-MT	GNMT	2.1.1.20	pig	I,II,III	D13308
	Glycine N-MT	GNMT	2.1.1.20	rat	I,II,III	X06150
	Guanidinoacetate N-MT	(GANMT)	2.1.1.2	rat	I,II,III	J03588
	Histamine N-MT	HNMT	2.1.1.8	rat	I,II,III	D10693
	Diphthamide N-MT	*DPH5*	2.1.1.98	yeast	nd	M83375
Small molecule S-MTase	Thioether S-MT	(TSMT)	2.1.1.96	mouse	I,II,III	M88694
Porphyrin precursor C-MTases	Precorrin-2 MT	*cobI*		*P. denitrificans*	I	M59301
	Precorrin-3 MT	*cobF*		*P. denitrificans*	I,III	M59301
	Precorrin-3 MT	*cobJ*		*P. denitrificans*	I,III	M59301
	Precorrin-3 MT	*cobL*		*P. denitrificans*	I,II,III	M59301
	Precorrin-3 MT	*cobM*		*P. denitrificans*	I,III	M59301
	Uroporphyrinogen III MT	*cobA*	2.1.1.107	*B. megaterium*	I,II,III	M62881
	Uroporphyrinogen III MT	UMT	2.1.1.107	*M. ivanovvii*	I,III	M62874
	Uroporphyrinogen III MT	UMT	2.1.1.107	*Pseudomonas sp.*	I,III	M32223
	Uroporphyrinogen III MT	*cysG*	2.1.1.107	*E. coli*	I,II,III	P11098
	Uroporphyrinogen III MT	*cysG*	2.1.1.107	*S. typhimurium*	I,II,III	P25924
	Magnesium protoporphyrin MT	bchH	2.1.1.11	*R. capsulatus*	nd	M74001
Lipid MTases	DHPB O-MT	COQ3	2.1.1.64	rat	I,II,III	L20427
	DHHB O-MT	*COQ3*	2.1.1.64	yeast	I,II,III	M73270
	UbiG O-MT	*ubiG*		*E. coli*	I,II,III	M87509
	Phosphatidylethanolamine MT	*PEM1*	2.1.1.17	yeast	nd	M16987
	Phosphatidylethanolamine MT	*pmtA*	2.1.1.71	*R. sphaeroides*	I,II,III	LO7247
	Phospholipid MT	*PEM2*	2.1.1.71	yeast	nd	M16988
	Cyclopropane fatty acid synthase	*cfa*	2.1.1.79	*E. coli*	I,II,III	M98330

[a] Nonstandard gene designations in parentheses.

[b] nd, not detected.

[c] Accession numbers are from the Genbank/EMBL release 77 (Roman type), NBRF PIR release 36 (italicized type), and Swissprot release 24 (underlined).

TABLE I—*Continued*

	Enzyme	Gene[a]	EC Number	Organism	Motifs[b]	Accession[c]
RNA MTases	tRNA C5 U MT	*trmA*	2.1.1.35	*E.coli*	**nd**	M57568
	tRNA N1 G MT	*trmD*	2.1.1.31	*E.coli*	II,III	X01818
	tRNA N2,N2 G MT	*trm1*	2.1.1.32	yeast	**nd**	M17193
	rRNA G MT	*grm*	2.1.1.51/52	*M. purpurea*	**nd**	M55520
	rRNA G MT	*grm*	2.1.1.51/52	*S. tenebrarius*	**nd**	*S17717*
	rRNA G MT	*grm*	2.1.1.51/52	*M. rosea*	**nd**	M55521
	rRNA MT	*sgm*		*S. zionensis*	**nd**	*S49806*
	rRNA MT	*lmrB*		*S. lincolnensis*	III	X62867
	rRNA MT	*kamB*		*S. tenebrarius*	**nd**	M64625
	rRNA MT	*kamC*		*S. hirsuta*	**nd**	M64626
	rRNA MT	*carb*		*S. thermotolerans*	I	M16503
	rRNA N6 A MT	*lrm*	2.1.1.48	*S. lividans*	I	*JS0635*
	rRNA N6 A MT	*ERMA*	2.1.1.48	*C. diphtheriae*	I,III	X51472
	rRNA N6 A MT	*ermA*	2.1.1.48	*S. aureus*	I,III	*A25101*
	rRNA N6 A MT	*ermBC*	2.1.1.48	*E. coli*	I,III	*B27739*
	rRNA N6 A MT	*ermBP*	2.1.1.48	*C. perfringens*	I,III	M77169
	rRNA N6 A MT	*ermC*	2.1.1.48	*S. aureus*	I	M19652
	rRNA N6 A MT	*ermCd*	2.1.1.48	*C. diphtheriae*	I	M36726
	rRNA N6 A MT	*ermD*	2.1.1.48	*B. licheniformis*	I	M29832
	rRNA N6 A MT	*ermE*	2.1.1.48	*S. erythaeus*	I	M11200
	rRNA N6 A MT	*ermF*	2.1.1.48	*B. fragilis*	I	*A25157*
	rRNA N6 A MT	*ermG*	2.1.1.48	*B. sphaericus*	I,III	M15332
	rRNA N6 A MT	*ermJ*	2.1.1.48	*B. anthracis*	I	L08389
	rRNA N6 A MT	*ermK*	2.1.1.48	*B. licheniformis*	I	*B42473*
	rRNA N6 A MT	*ermM*	2.1.1.48	*S. epidermidis*	I,III	*A24497*
	rRNA N6 A MT	*emrR*	2.1.1.48	*L. reuteri*	I,III	M64090
	rRNA N6 A MT	*ermSF*	2.1.1.48	*S. fraediae*	I,III	M19269
	rRNA N6,N6 A MT	*ksgA*		*E. coli*	I	X06536

ferase, and in AdoHcy hydrolase (7). Other studies have added examples to motif I (8–12) and motif III (11, 12).

To more clearly establish the limits of using these motifs to characterize methyltransferases we have now extended our analysis of methyltransferase sequences. We find that 45 of 84 available methyltransferase sequences contain all three motifs spaced at similar intervals, and an additional 15 enzymes possess motifs I and III but appear to lack motif II. We suggest that these sequence motifs may represent core elements of the polypeptide that are brought together in the three-dimensional structure to interact directly with this cofactor. Thirteen of the 84 enzymes do not appear to contain any of these motifs. These latter enzymes may have simply diverged too greatly to identify the motifs or may represent an independent approach to AdoMet or AdoHcy utilization. We have also discerned intriguing sequence similarities between motifs I and III in methyltransferases, AdoMet synthetases, AdoHcy hydrolases, and AdoMet decarboxylases. We speculate that these similar motifs may comprise part of an evolutionarily conserved AdoMet and AdoHcy binding structure in these classes of enzymes.

METHODS

Multiple sequence alignments were carried out using the Megalign program from DNASTAR (Madison, WI). Multiple alignments are created according to the CLUSTAL V method (13). The initial parameters used to construct the alignments were as follows: a k-tuple of 1 or 2, a window of 5 residues, and a gap penalty of 3 were applied to the pairwise alignments. A gap penalty of 10 and a gap length penalty of 10 or 20 were applied to the multiple alignments. All alignments were evaluated using the PAM 250 distance matrix (14). Two-way protein sequence alignments were carried out in Megalign acording to the Lipman–Pearson method (15).

The consensi for each motif was determined by computing the amino acid frequency at each position as described in the legend to Fig. 1. The frequencies were divided by the Dayhoff frequency for each amino acid (14) and amino acids that scored 2.5 or greater were chosen for the consensus. We did not apply this standard to tryptophan residues because the Dayhoff frequency of tryptophan is only 1% and a single tryptophan residue in a group of 26–29 distinct enzymes would score above 2.5 by this method.

RESULTS

Excluding motif I, most DNA methyltransferases do not appear to have conserved sequence blocks in common with protein, lipid, and small-molecule C-, N-, O-, and S-methyltransferases. As the conserved sequence motifs of DNA methyltransferases have been fully described elsewhere (4, 5, 16), we chose to focus here on the sequence motifs in enzymes active on substrates other than DNA (Table I). We selected the 84 sequences presented in Table I from searches of the translated GenBank database re-

TABLE II

Methyltransferase Sequence Motifs I–III

Motif I		Motif II		Motif III		Gene	Organism
	L IVEV C P VLDIGGGTG		A Y L GTY VIV PQFDAIFC		V L I K IIFL LLRPGGRLLI	Consensus[a]	
81	ALDVGSGSG	150	APYDAIHV	171	QLKPGGRLIL	PIMT	human
81	ALDVGSGSG	150	APYDAIHV	171	QLKPGGRLIL	PIMT	bovine
82	ALDVGSGSG	151	APYDAIHV	171	QLKPGGRLIL	PIMT	mouse
82	APDVGSGSG	151	APYDAIHV	172	QLKPGGRLIL	PIMT	rat
87	ALDVGSGSG	157	APYDAIHV	178	QLKPGGRMVI	PIMT	wheat
79	VLEIGTGSG	140	APFDAIIV	161	QLDEGGILVL	*pcm*	*E. coli*
138	ADTLGTAPG	222	GPFDAIFC	251	LLKPDGLLFA	*cheR*	*E. coli*
138	ADALGMAPG	222	GPFDAIFC	251	LLKPDGLLFA	*cheR*	*S. typhimurium*
127	LAELGALSL	209	SSLDLILC	238	ALRPGGLLFL	*frzF*	*M. xanthus*
	b	271	PEADLYIL	300	TCKPGGGILV	HOMT	human
183	ICDLGGGSG	243	PEADLYIL	272	ACRTGGGILV	HOMT	bovine
184	IYDLGGGGG	244	PEADLYIL	273	ACRPGGGVLL	HOMT	chicken
112	LLELGAYCG	183	DILDMVFL	209	LLRKGTVLLA	COMT	human
62	VLELGAYCG	133	DILDMVFL	159	LLRKGTVLLA	COMT	rat
205	LVDVGGGVG	258	PAGDAILM	288	LPENGKVIVV	(CAOMT)	maize
204	LVDVGGGIG	257	PKADAVFM	287	LPDNGKVIVA	(CAOMT)	alfalfa
204	LVDVGGGIG	257	PKADAVFM	287	LPENGKVILV	(CAOMT)	aspen
77	TMEIGVYIG	149	GIFDFVFV	173	LVKIGGLIGY	(CCOMT)	parsley
208	VVDIGGADG	269	GGGDLYVL	298	AMPAHARLLV	*dmpM*	*S. alboniger*
232	VMDVGGGIG	288	QGADVITL	317	ALPPGGRLII	*crtF*	*R. capsulatus*
205	LVDVGGNIG	258	PQADAIFM	287	SLAKGGKIIL	*imt1*	*M. crystallinum*
183	VLDVGGGKG	244	RKADAIIL	273	ALEPGGRILI	*dnrK*	*S. peucetius*
331	IADLGGGDG	393	TGYDAYLF	423	IGDDDARLLI	*tcmN*	*S. glaucesens*
173	FVDLGGARG	234	PRADVFIV	263	ALTPGGAVLV	*tcmO*	*S. glaucesens*
64	VLEIGTFIG	135	GAFDLVFV	159	LVRPGGLVAI	*mdmC*	*S. mycarofaciens*
85	VLDVGFGLG	149	EIFDRVTS	176	VLKPGGVLAI	*eryG*	*S. erythraea*
75	LIDIGSGPT	173	LPADALVS	204	LLRPGGHLLL	PNMT	human
75	LIDIGSGPT	173	LPADALVS	204	LLRPGGHLLL	PNMT	bovine
51	LIDIGSGPT	149	LPADALVS	180	LLRPGGHLLL	PNMT	rat
86	LIDIGSGPT	184	LPADALVS	215	LLRPGGHLLL	PNMT	mouse
59	VLDVACGIG	129	GGFDAVIC	164	MVRTGGLLVI	GNMT	rabbit
58	VLDVACGIG	128	GGFDAVIC	163	MVRSGGLLVI	GNMT	pig
61	VLDVACGIG	129	DGFDAVIC	164	MVRPGGLLVI	GNMT	rat
64	VLEVGFGMA	127	GHFDGILY	159	LLKPGGILTY	(GANMT)	rat
56	ILSIGGGAG	135	PKWDFIHM	156	LKFFHGLLAA	HNMT	rat
60	LIDIGSGPT	156	PLADCVLT	187	LLKPGGHLVT	(TSMT)	mouse
7	GRLIGVGIG			112	AFLVWGDPML	*cobI*	*P. denitrificans*
9	ILIIGIGSG			106	AVLSEGDPLF	*cobF*	*P. denitrificans*
5	LYVVGTGPG			76	CMVSGGDPGV	*cobJ*	*P. denitrificans*
265	LWDIGGGSG	328	PQPDAIFI	351	ALKSGGRLVA	*cobL*	*P. denitrificans*
3	VHFIGAGPG			77	ARLHSGDLSV	*cobM*	*P. denitrificans*
4	VYLVGAGPG	47	SDADIIYC	83	TRLKGGDPEV	*cobA*	*B. megaterium*
3	VYLVGAGPG			81	VRLKGGDPEV	UMT	*M. ivanovvii*
17	VWLVGAGPG			96	LRLKGGDPEV	UMT	*Pseudomonas sp.*
218	VVLVGAGPG	261	RDADRVFV	297	VRLKGGDPFI	*cysG*	*E. coli*
218	VVLVGAGPG	261	RDADRVFV	297	VRLKGGDPFI	*cysG*	*S. typhimurium*
91	ILDVGCGGG	155	ECFDAVVA	182	VLKPGGSLFI	COQ3	rat
130	VLDVGCGGG	192	GQFDIITC	220	LNPEKGILFL	*COQ3*	yeast
76	VLDVGCGGG	137	GQYDVVTC	164	LVKPGGDVFF	*ubiG*	*E. coli*
42	VLEVGVGIG	103	EIFDTVVA	130	VCRKGGEVVI	*pmtA*	*R. sphaeroides*
171	VLDIGCGWG	227	DQFDRIVS	256	NLKPEGIFLL	*cfa*	*E. coli*
		66	PLRDAIHA	101	ELATNQKLIL	*trmD*	*E.coli*
				167	PSVDGGILVI	*lmrB*	*S. lincolnensis*

[a] The consensus at each position is defined as the residue(s) present at a frequency of 2.5 times or more of the natural abundance for each amino acid. The first consensus row immediately above the sequences represents the most common choice for that position above the cutoff. The second, third, and fourth most common residues at each position above the cutoff are in the rows above.

[b] Human HOMT has a 26-residue LINE 1 insertion in motif I, and it was not included in this table.

TABLE II—*Continued*

Motif I	Motif II	Motif III	Gene	Organism
78 VLEVGAGNG			*carb*	*S. thermotolerans*
39 LLEVGAGNG			*lrm*	*S. lividans*
78 VLEVGAGNG		169 PNVDGGILVI	*ERMA*	*C. diphtheriae*
36 IIEIGPGSG		167 PSVDSVLIVL	*ermA*	*S. aureus*
34 VIEIGSGKG		166 PRVNSVLIKL	*ermBC*	*E. coli*
33 VYEIGTGKG		166 PKVNSVLIKL	*ermBP*	*C. perfringens*
33 VYEIGTGKG			*ermC*	*S. aureus*
36 IIEIGPGSG			*ermCd*	*C. diphtheriae*
35 IFEIGSGKG			*ermD*	*B. licheniformis*
48 VLELGAGKG			*ermE*	*S. erythaeus*
65 VLEAGPGEG			*ermF*	*B. fragilis*
37 VLDIGAGKG		167 PKVDSALIVL	*ermG*	*B. sphaericus*
55 VLELGAGKG			*ermJ*	*B. anthracis*
48 VLELGAGKG			*ermK*	*B. licheniformis*
34 IFEIGSGKG		167 PKVNSSLIRL	*ermM*	*S. epidermidis*
53 IIEIGSGKG		186 PRVNSSLIVL	*emrR*	*L. reuteri*
89 LLEVGAGRG		226 PRVDSGILRI	*ermSF*	*S. fraediae*
41 MVEIGPGLA			*ksgA*	*E. coli*

leases 76 and 77, the NBRF-PIR database version 36, and from Swissprot release 24 using the keywords "methyltransferase" or "methylase." We eliminated duplicate entries, partial sequences, and putative sequences. These sequences represent 37 distinct enzymatic activities; 3 protein carboxyl methyltransferases, 12 small-molecule O-methyltransferases, 5-small molecule N-methyltransferases, 1 small-molecule S-methyltransferase, 4 porphyrin precursor methyltransferases, 5 lipid methyltransferases, and 7 RNA methyltransferases.

Methyltransferase motif I. Methyltransferase motif I has been previously described in DNA methyltransferases (4, 5, 16, 17) as well as in a variety of bacterial RNA methyltransferases and in procaryotic and eucaryotic small-molecule and protein methyltransferases (7–12). We have now extended our analysis to an expanded sequence database of 84 methyltransferases and found this region to occur in 69 sequences representing 29 distinct enzymes (Tables I and II). We define motif I as a nine-residue block with the consensus sequence (V/I/L)(L/V)(D/E)(V/I)G(G/C)G(T/P)G as shown in Table II and Fig. 1A. The amino acids at each position of the consensus were found to be present at a frequency of 2.5 times or more than that of their Dayhoff frequency (see Methods). The glycine residue at position 5 is present in all of the sequences except the glycine methyltransferases where it is substituted by an alanine residue. The glycines at positions 7 and 9 in this motif are present in 59 and 62 of the sequences, respectively. This consensus is in agreement with the general consensus given previously for motif I, hh(D/E)hGXGXG, where "h" represents a hydrophobic residue (9). The DNA methyltransferases lack the glycine at position 5 and often have a phenylalanine at this position (6). The consensus for this region in a number of m^5C and m^6A DNA methyltransferases has been defined as hh(D/S)(L/P)FXGXG (6).

The protein γ-glutamyl carboxyl methyltransferases from *Escherichia coli* and *Salmonella typhimurium* have the poorest match to the consensus for motif I. They are missing an acidic residue at position 3 and they match the consensus only at the glycine residues in positions 5 and 9 and at the proline residue in position 8. However, these methyltransferases have well-defined motifs II and III (Table II). The bacterial precorrin-3 and uroporphyrinogen methyltransferases, with the exception of CobL, also lack the characteristic acidic residue at position 3. Instead, a hydrophobic residue is found at this position (Table II).

Motif I is followed in 67 of 69 cases by an aspartate or a glutamate residue 17–19 residues C-terminal to the motif. The exceptions to this are the rat histamine methyltransferase which has an asparagine at this position and the *Pseudomonas denitrificans* precorrin-3 methyltransferase CobM, which has a cysteine at this position. This acidic residue is frequently preceded by a number of hydrophobic residues. The consensus is hhXh(D/E), where "h" is a hydrophobic residue (data not shown).

Methyltransferase motif II. As shown in Table II and Fig. 1B, motif II comprises an eight-residue conserved region that is found 57 ± 13 SD (range of 36–90) residues after the glycine delineating the end of motif I (Fig. 2A). It is present in 46 of the 84 sequences analyzed (26 distinct enzymes). The consensus sequence for this region is (P/G)(Q/T)(F/Y/A)DA(I/V/Y)(F/I)(C/V/L) (Table II and Fig. 1B). The central aspartate is invariant. There are a number of positions in close proximity to motif II that

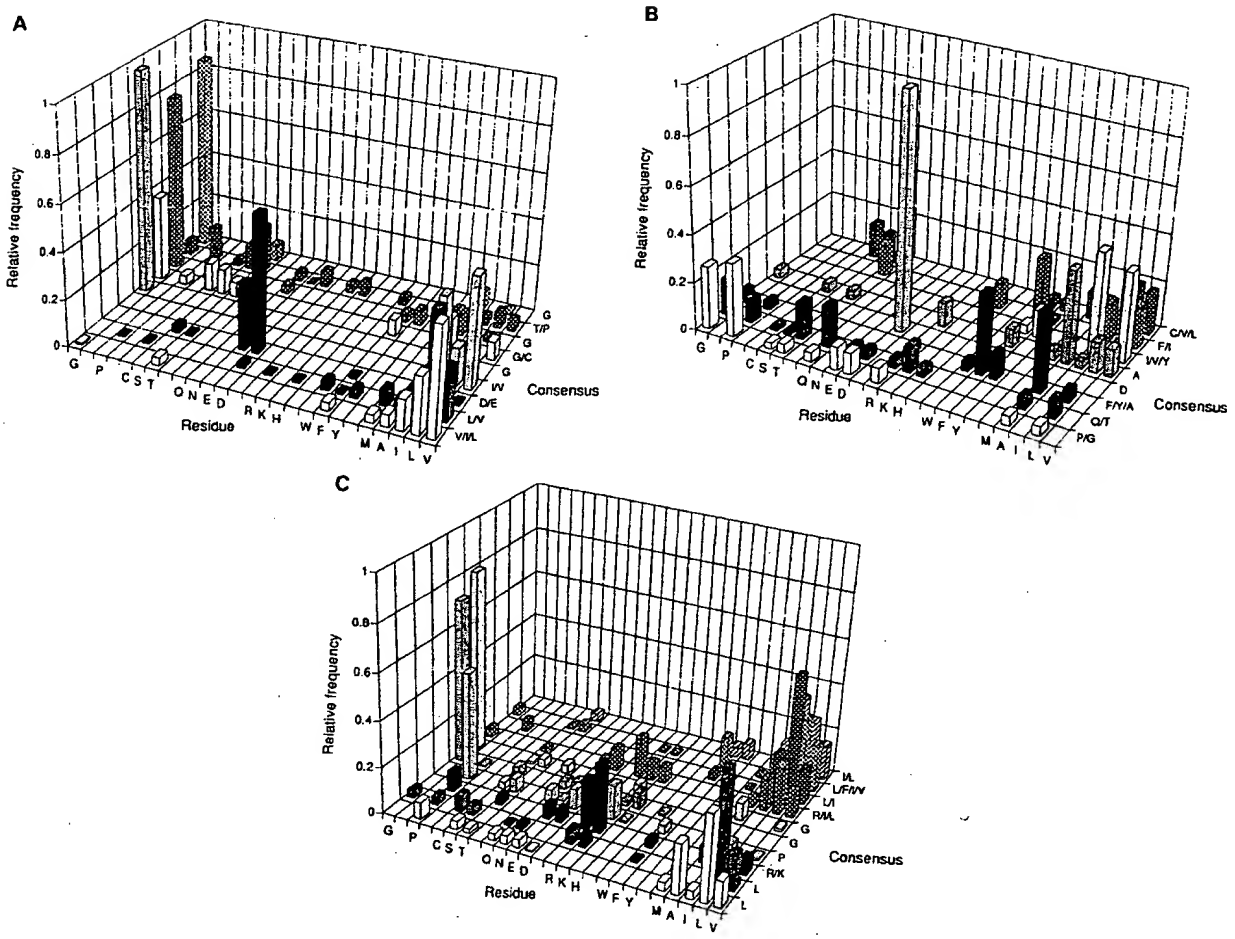


FIG. 1. Relative amino acid distribution in methyltransferase sequence motifs. (A) Motif I. (B) Motif II. (C) Motif III. Individual enzymes were assigned a fractional weighting factor inversely proportional to the number of instances that the distinct enzymatic activity occurred in different species in Table I. For example, there are three glycine N-methyltransferases in Table I, so they each received a weighting factor of 0.333. The residues at each position were tabulated in one column and their respective weighting factors were tabulated in the adjacent column. The residues and the weighting factors were then sorted by amino acid and the weighting factors for each amino acid were then summed up to give the frequency of occurrence of each amino acid at each position. The relative frequency at each position was calculated by dividing the amino acid frequency by the number of distinct enzymatic activities in Table I.

are unusually rich in the aromatic amino acids phenylalanine, tyrosine, and tryptophan. Figure 3 shows that at positions −8, −1, and +3 with respect to the central aspartate, the relative frequencies of aromatic residues are 0.31, 0.49, and 0.29, respectively. In contrast, the Dayhoff frequency for the three aromatic residues totals only 0.08 (12). Of 26 distinct enzymes that have region II, only one, the TrmD tRNA methyltransferase in *E. coli,* is devoid of aromatic residues in this span.

Methyltransferase motif III. As shown in Table II and Fig. 1C, motif III comprises a 10-residue conserved region that is found 22 ± 5 SD (range of 12–38) residues after the end of motif II (Fig. 2B). It is present in 61 of the 84 sequences examined (28 distinct enzymes). The consensus sequence for this region is LL(R/K)PGG(R/I/L)(L/I)(L/F/I/V)(I/L). (Table II and Fig. 1C). The central glycines of this region are highly conserved and at least one of them is found in all sequences except seven of the RNA methyltransferases and the *O*-demethyl puromycin methyltransferase in *Streptomyces alboniger.*

Methyltransferase sequence motifs in other AdoMet or AdoHcy-utilizing enzymes. We then examined the sequences of other enzymes that utilize the substrate or the product of methyltransferases, AdoMet or AdoHcy, to ascertain whether they, too, possessed motifs I, II, and III. AdoMet decarboxylase (EC 4.1.1.50) in *E. coli* possesses all three sequence motifs in the same relative positions as in the methyltransferases. However, the three known

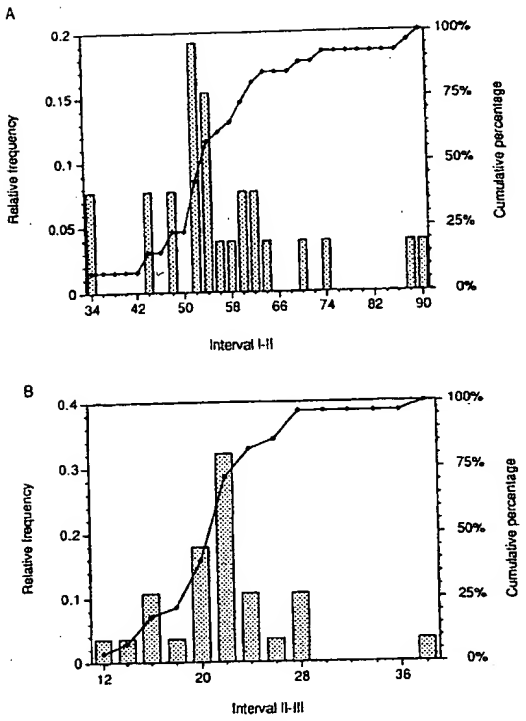


FIG. 2. Distance between methyltransferase sequence motifs. The number of residues between motifs I and II or between motifs II and III for each enzyme was tabulated. A single average value was calculated for each of the 37 distinct types of enzymes in Table I. The frequency of occurrence at two-residue intervals was calculated and tabulated (e.g., "34" = 34–36, etc.). The relative frequency was calculated by dividing the frequency for each observation by the sum of the frequencies. (A) Distances between motifs I and II in intervals of two residues. (B) Distances between motifs II and III in intervals of two residues.

mammalian AdoMet decarboxylases possess only motif III (Table III), while the yeast enzyme does not appear to have any of these motifs. We then examined the sequences of the AdoHcy hydrolases (EC 3.3.1.1) for the presence of methyltransferase sequence motifs. We found that all seven AdoHcy hydrolase sequences possess the three motifs (Table III). The interval between motif I and motif II is 65 residues in all cases and the interval between motif II and motif III is 33 ± 1 SD residues. The first interval is comparable to that in methyltransferases (57 ± 13 SD) and the second interval is somewhat larger than that found for the methyltransferases (22 ± 5 SD).

The AdoMet synthetases (EC 2.5.1.6) possess only motifs I and III, separated by 107 residues. The consensus for motif I in both the AdoMet synthetases and the AdoMet hydrolases, however, is GXGDXG. This corresponds to the ATP-binding motif in protein kinases, GXGXXG, that is also found in other nucleotide binding proteins such as NAD-binding proteins, ras-like rho proteins, and in ATP synthase β and adenylate kinase (18, 19). In the AdoMet synthetases and the AdoHcy hydrolases, motif III has only one of the two central glycines found in this motif in the methyltransferases. However, the motif does end in four generally hydrophobic residues (Table III).

Predictive potential of methyltransferase sequence motifs. Computerized searches of the protein sequence databases may be used to detect sequence similarities between proteins of known function and uncharacterized ORFs. However, search queries of the database with full-length protein sequences may fail to detect database entries that have only noncontiguous, short homologies to the query sequence. A search of the NBRF-PIR release 36, translated GenBank release 77, and Swissprot release 25 databases using the consensi we defined for either motifs I or II or III and allowing two, one, and two mismatches, respectively, retrieved 424 entries of a total of 73,582. After elimination of duplicate entries, 353 entries, or 0.48% of the total database, remained. Of these, 37 were methyltransferases representing 21 distinct enzymatic activities that were described among the 68 enzymes in Table I that possess at least one of the motifs and are included in the versions of the database indicated above (data not shown). An additional three DNA methyltransferases that were not included in this study were also retrieved. This search also retrieved 30 hypothetical proteins or ORFs. Of these, 27 were retrieved with motif I, five were retrieved with motif III, and two were retrieved with both motifs I and III. Visual inspection of these sequences found seven entries that had two or three of the methyltransferase motifs in the expected order and spacing (Table IV). None of the retrieved proteins, other than methyltransferases or putative methyltransferases, were found to contain more than one of the three sequence motifs. Most of the ORFs in Table IV show sequence similarities of 20% or greater over 200 or more residues to one or more known methyltransferases, further supporting the conclusion that these ORFs encode methyltransferases. The ORFs JS0718, IN37_SPIOL, YYAP_YEAST, and YCPW_PSEA9 show a greater degree of sequence similarity to a group of related small-molecule *O*-methyltransferases, UbiG, COQ3, and EryG than to other methyltransferases (Table IV). While this similarity may be suggestive, no sequence blocks that are diagnostic of this particular subgroup of *O*-methyltransferases have been identified, so inferences regarding the substrate specificity of these ORFs can only be tentative.

A more stringent consensus search, allowing only a single mismatch in each motif, retrieved only 102 entries, of which 25 were methyltransferases, representing 13 distinct enzymatic activities, three were putative methyltransferases, and 75 were unrelated proteins possessing a single consensus motif (data not shown).

DISCUSSION

Previous work has shown that many AdoMet and AdoHcy utilizing enzymes, including methyltransferases,

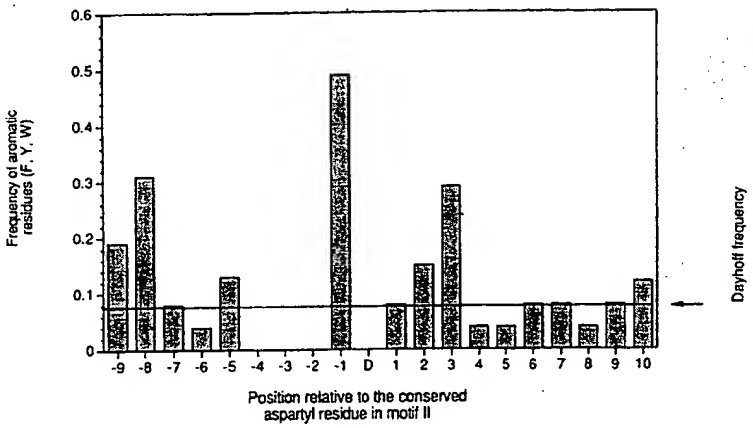


FIG. 3. Distribution of the aromatic amino acids phenylalanine (F), tyrosine (Y), and tryptophan (W) in and around methyltransferase motif II. The number of aromatic residues at each position 9 residues N-terminal and 10 residues C-terminal to the central aspartate of motif II was tabulated. Individual enzymes were assigned a fractional weighting factor as described in Fig. 1. The weighted sum of aromatic residues at each position was then divided by 26, the number of distinct enzymatic activities possessing motif II. Dayhoff frequency, combined natural abundance of F, Y, and W (14).

AdoMet synthetases, AdoMet decarboxylases, and AdoHcy hydrolases, have a common sequence element termed here motif I (4, 5, 7–12). In this work, we have shown that motif I has the consensus (V/I/L)(L/V)(D/E)(I/V)G(G/C)G(T/P)G and is present in most methyltransferases sequenced to date, including 69 non-DNA methyltransferases representing 29 distinct enzymes. A more limited consensus, hh(D/S)(L/P)FXGXG, was identified in DNA cytosine-specific and adenine-specific methyltransferases (6). In these enzymes, the phenylalanine and the two glycines are the most conserved. However, the other conserved residues may also be found in some of these sequences. For example, the *Pst*I adenine methyltransferase has the sequence ILDAGAGVG, which matches the more general methyltransferase motif I consensus hh(D/E)hGXGXG, where "h" is a hydrophobic residue (9). The recent determination of the crystal structure of the *Hha*I DNA m^5C methyltransferase has confirmed the involvement of motif I in AdoMet binding (17). Motif I comprises part of the AdoMet binding pocket of this enzyme and the conserved phenylalanine, along with other hydrophobic residues, forms a hydrophobic platform on one side of the purine and ribose rings of AdoMet. The glycines in this motif form a tight loop responsible for properly positioning the AdoMet in the binding pocket (17).

The sequence heterogeneity of this region has, on occasion, made this motif difficult to detect. We were able to make use of multiple alignments of clusters of related methyltransferases to detect the presence of motif I in two methyltransferase sequences that had previously been reported to lack motif I. These are the midamycin O-methyltransferase (mdmC; Ref. 20) and the caffeoyl-CoA 3-O methyltransferase (21). This method has the advantage of being able to align sequences of interest with similar sequences that have well-defined methyltransferase motifs. It is often possible to identify motifs by visual inspection in such an expanded context.

Two other regions, termed motifs II and III, were originally described in a smaller number of methyltransferases (7). Other investigators have not always identified these motifs in recently sequenced methyltransferases and hence some have questioned the general significance of these regions (8–12). However, we have been able to use multiple alignments to show that motifs II and III are present in both UbiG (9) and in EryG (8), for example. We found that motif II is present at a distinct interval of 57 ± 13 SD C-terminal to motif I and is present in nearly as many distinct enzymes as motif I (26 for the former, 29 for the latter). The main exceptions to this rule are the RNA methyltransferases and a number of the porphyrin precursor methyltransferases (Table I). Motif II is also found in the AdoHcy hydrolases and in the *E. coli* AdoMet decarboxylase, positioned 65 and 60 residues, respectively, from the end of motif I (Table III). However, in the AdoHcy hydrolases, only positions 4–8 are conserved with respect to the sequences found in methyltransferases.

Motif II is abundant in aromatic residues. At positions −8, −1, and +3 with respect to the central aspartate, there are respectively 3.9, 6.1, and 3.6 times more phenylalanine, tyrosine, and tryptophan residues than expected from the combined Dayhoff frequencies for these residues (Fig. 3). The possibility that the aromatic residues in, or in close proximity to motif II, may be involved in binding AdoMet is supported by recent work showing that positively charged quaternary ammonium, quinolinium, and sulfonium compounds can be bound by aromatic groups in

TABLE III

Motifs I–III in AdoMet Decarboxylases, AdoHcy Hydrolases, AdoMet Synthetases, and NTP Binding Proteins

motif I	motif II	motif III	
L IVEV C P VLDIGGGTG	GTY VIV PQFDAIFC	V L I K IIFL LLRPGGRLLI	Methyltransferase consensus
		DLIPGSVIDAT	AdoMet decarboxylase majority
81 VLIIGGGDG	150 QTFDVIIS	181 CLNPGGIFVAQ	AdoMet decarboxylase (*E. coli*)
		211 DLIPGSVIDAT	AdoMet decarboxylase (human)
		211 DLIPGSVIDAT	AdoMet decarboxylase (rat)
		211 DLIPGSVIDAT	AdoMet decarboxylase (hamster)
hhhhGYGDVGK[a]	MbDDAIVC[a]	LKNGRbhhhh[a]	AdoHcy hydrolase majority
216 AVVAGYGDVGK	290 MKDDAIVC	329 RLKNGRRIILL	AdoHcy hydrolase (human)
216 AVVAGYGDVGK	290 MKDDAIVC	329 LLKNGHRIILL	AdoHcy hydrolase (rat)
218 AVVAGYGDVGK	292 LPNDAIVC	331 TLKNGRHVILL	AdoHcy hydrolase (*C. elegans*)
265 ALIAGYGDVGK	339 MKNNAIVC	380 FPDTGRGIIIL	AdoHcy hydrolase (parsley)
215 CCVCGYGDVGK	289 MRDDAIVC	328 TMENGRHIILL	AdoHcy hydrolase (Leismania)
215 AVVAGYGDVGK	289 MKEDAIVC	327 TLANGVHIILL	AdoHcy hydrolase (mold)
265 AVVCGYGDVGK	339 MKNNAIVC	380 FPETKTGIIVL	AdoHcy hydrolase (*T. aestivum*)
EaIGAGDQG		LNPSGRFVIG	AdoMet synthetase majority
127 EEDIGAGDQGL		243 HLQPSGRFVIG	AdoMet synthetase (human)
115 PEDIGAGDQGH		231 HLNPSGRFVIG	AdoMet synthetase (Arabidopsis)
116 PEEIGAGDQGH		232 HLNPSGRFVIG	AdoMet synthetase (poplar)
116 LEDLGAGDQGI		232 FIQPSGRFVIG	AdoMet synthetase (yeast)
113 PLEQGAGDQGL		225 FINPTGRFVIG	AdoMet synthetase (*E. coli*)

[a] 'a', DE; 'h', MAILV; 'o', FYW; 'b', HKR.

cation–π interactions that occur between the positively charged group and the π electrons of the aromatic ring (22, 23). For example, an artificial host molecule rich in aromatic rings can bind acetylcholine with a K_d of 50 μM (22, 23). [^{3}H]acetylcholine mustard, a reactive analog of acetylcholine, was found to label the *Torpedo* nicotinic acetylcholine receptor at Tyr-93 on the α-subunit in the conserved region WXPDhhhYN, where "h" is a hydrophobic residue, at the proposed binding site of acetylcholine (24). Recent crystallographic and affinity labeling studies of binding sites of the *Torpedo* acetylcholinesterase for quaternary ligands have demonstrated that the quaternary groups interact with the indole rings of Trp-84 and Trp-279 in this enzyme's active site (25).

Further chemical studies of cation–π interactions found that an aromatic host molecule can catalyze methyl group transfer from an aryldimethyl sulfonium compound to thiocyanate to form methylthiocyanate (23). It was hypothesized that catalysis is effected by stabilizing the positively charged transition state through cation–π interactions. These authors suggest that the positively charged sulfonium on AdoMet may be bound and stabilized by methyltransferases through the same type of interaction, and predict that the active sites of methyltransferases may be rich in aromatic residues.

Motif III is found at an interval of 22 ± 5 SD residues C-terminal to motif II in 28 distinct enzymes. It is absent only in a number of RNA methyltransferases (Tables I and II). It is also found in AdoMet decarboxylases, AdoMet synthetases, and AdoHcy hydrolases. However, usually only one of the two conserved central glycines is found in these latter enzymes. Site-directed mutagenesis of motif III of the rat guanidinoacetate N-methyltransferase at five individual positions did not markedly alter the K_m for AdoMet, or the catalytic activity of this enzyme, although a twofold range of K_i values for AdoHcy was observed (11). These investigators thus question whether this region forms part of the active site (11, 12). However, the 10-residue motif III of this enzyme would still match the consensus sequence for this motif in 9 of

TABLE IV

Methyltranferase Motifs in Hypothetical ORFs in the Sequence Databases

Database and identifier[a]		Motif I		Motif II		Motif III	Δ1-2[b]	Δ2-3[c]	Database Homologies[d]	Consensus matches
Methyltransferase consensus:		L IVEV C P VLDIGGGTG		A Y L GTY VIV PQFDAIFC		V L IV K IIFL LLRPGGRLLI	59	22		
E. coli: YIGO_ECOLI (SW)	67	VLDLAGGTG	132	NTFDCITI	159	VLKPGGRLLV	57	20	PIMT (human): (24.2%, 145 aa); PrmA: (23.9%, 200 aa)	I, III
E. coli: JS0718 (PIR)[e]			60	NTFDIVIS	87	ILKPGGRLIV		20	EryG (21.7%, 174 aa)	III
S. fradiae: YT37_STRFR (SW)	130	ALDLGCGPG	192	GSIDCART	219	VLRPGGRLVM	54	20	PIMT (bovine): (24.5%, 183 aa)	I, III
Pseudanabena: YCPW_PSEA9 (SW)[f]	57	ILDAGCGSG							COQ3 (yeast): (30.8%, 86 aa)	I
Yeast: L12000 (GB)	168	VLELGCGKG	241	FPCDIVST	272	SLKIGGHFFG	65	24	PNMT (rat): (16.5%, 256 aa)	I
Yeast: YYAP_YEAST (SW)	123	VLDVGCGVG	186	NTFDKVYA	213	VLKPGGTLVM	55	20	pmtA: (25.4%, 131 aa); EryG: (24.3%, 233 aa); UbiG: (23.6%, 178 aa)	I
Yeast: YCT7_YEAST (SW)	51	ILDIGCGSG	109	GSFDAAIS	147	LKKGGKFVAQ	50	31	GNMT (rabbit): (24.6%, 232 aa)	I
Spinach: IN37_SPIOL (SW)	121	VVDVGGGTG	181	DYADRYVS	208	VLKLGGKACL	52	20	UbiG: (26.1%, 129 aa); EryG: (21.2%, 293 aa)	I

[a] PIR, NBRF-PIR release 36; GB, Genbank release 77; SW, Swissprot release 25.
[b] Interval between motifs I and II.
[c] Interval between motifs II and III.
[d] Highest similarity scores in a Lipman–Pearson alignment of the ORFs to the methyltransferases in Table I or to other ORFs in this table. The alignment parameters were ktup = 2, gap penalty = 4 (or 2, for YCPW_PSEA9 vs COQ3), and gap length penalty = 6.
[e] Partial C-terminal sequence.
[f] Partial N-terminal sequence.

10 positions, which is also the case in a variety of other methyltransferases (Table II).

There are three lines of evidence that suggest that the three motifs are involved in binding AdoMet or AdoHcy. First, the crystal structure of the *Hha*I m^{5}C DNA methyltransferase complexed with AdoMet discussed above has directly demonstrated that motif I is involved in AdoMet binding. Second, they are present in a large number of methyltransferases with diverse functions. As AdoMet and AdoHcy binding is the common feature of these diverse enzymes, the presence of three distinctly spaced sequence motifs suggests evolutionary conservation of protein regions involved in this activity. Third, an experiment aimed at directly characterizing the interaction of AdoMet with methyltransferases has found evidence for the interaction of residues in the vicinity of motif II with AdoMet. Tyr-136, located three residues C-terminal to motif II of the rat guanidinoacetate N-methyltransferase (EC 2.1.1.2), is exclusively photolabeled with S-adenosyl[^{3}H-*methyl*]methionine (26). In addition, the competitive inhibitors AdoHcy and sinefungin were able to block effectively the photoincorporation of radioactivity. This finding is consistent with the involvement of aromatic residues in this region in AdoMet binding discussed earlier.

There are, however, a number of methyltransferases that apparently do not possess any of these three motifs or additional elements of sequence similarity to other methyltransferases. These enzymes include the isoprenylcysteine protein carboxyl methyltransferase, the diphthamide biosynthesis N-methyltransferase, the phospholipid methyltransferases PEM1 and PEM2, and the bacterial rRNA methyltransferases Sgm and Grm (see Table I). These results suggest convergent evolution of AdoMet-dependent methylation. With the possible exception of the STE14 isoprenyl cysteine protein carboxyl methyltransferase and the PEM2 phospholipid methyltransferase, there are no obvious sequence similarities among the methyltransferases that lack the three sequence motifs discussed here nor between these methyltransferases and other proteins in the database. It is possible, therefore, that many of these enzymes may represent independent approaches to implementing AdoMet-dependent methylation. Likewise, crystallographic studies of the AdoMet-binding MetJ repressor in *E. coli*

show an independent approach to AdoMet binding. In this protein, AdoMet is bound by sequence segments unrelated to the motifs discussed in this paper and is found at the interface between the protein and the MetJ-binding DNA sequence (27).

As the number of unidentified ORFs in the databases continues to increase, the use of these methyltransferase sequence motifs coupled with a knowledge of their expected spacing may be useful in discerning candidate methyltransferases among these new sequences. Conventional sequence homology searches often fail to detect relationships between distantly related proteins because the overall sequence similarity is low and similarities may only be found in noncontiguous, short regions of local similarity. A query of the combined NBRF PIR release 36, translated GenBank release 77, and Swiss Prot release 25 databases using the consensus motifs defined here was able to pick out 37 methyltransferases that also appear in Table I. As only one of these (caffeoyl CoA O-MT) was identified exclusively with motif II, it appears that motifs I and III are the most useful for database searching. Although the consensus search allowing for two mismatches in motifs I and III and one mismatch in motif II also yielded approximately 300 false positives, these can be readily distinguished from methyltransferase candidates as only a single motif is found upon further visual inspection. Because there is significant heterogeneity in the methyltransferase sequence motifs the consensi that we have defined did not detect some of the methyltransferases listed in Table I. While it is possible to increase the number of hits by less stringently defining the consensi, this also increases the number of false positives. A more fruitful approach may be to employ less stringent consensus sequences and a search algorithm that is able to screen for the ordering and the spacing between the motifs. Work in this direction is currently underway in our laboratory.

Our knowledge of methyltransferase sequences is still fragmentary. There are 103 distinct AdoMet-dependent methyltransferases that have been assigned EC numbers (1) and at least another 11 distinct enzymes listed in Table I that have not yet been assigned EC numbers. For example, while several sequences of protein carboxyl methyltransferases are known (see Table I), no sequences are yet available for protein arginine, histidine, lysine, or N-terminal α-amino methyltransferases (28). The 37 distinct methyltransferases whose sequences were analyzed in this study thus comprise less than one-third of these enzymes. Sequence determinations of more of these enzymes, as well as of yet-to-be-discovered methyltransferases, may be of use in discerning additional blocks of sequence similarity between enzymes of related function to understand better the evolutionary relationships between the various methyltransferases.

ACKNOWLEDGMENTS

This work was supported by Grant GM-26020 from the National Institutes of Health to S.C. R.M.K. was supported in part by United States Public Health Service Training Grant GM-07185.

REFERENCES

1. Webb, E. C. (1992) Enzyme Nomenclature, Academic Press, San Diego.
2. Dever, T. E., Glynias, M. J., and Merrick, W. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1814–1818.
3. Kaziro, Y., Itoh, H., Kozasa, T., Nakafuku, M., and Satoh, T. (1991) *Annu. Rev. Biochem.* **60**, 349–400.
4. Lauster, R., Trautner, T. A., and Noyer, W. M. (1989) *J. Mol. Biol.* **206**, 305–312.
5. Posfai, J., Bhagwat, A. S., Posfai, G., and Roberts, R. J. (1989) *Nucleic Acids Res.* **17**, 2421–2435.
6. Lauster, R. (1989) *J. Mol. Biol.* **206**, 313–321.
7. Ingrosso, D., Fowler, A. V., Bleibaum, J., and Clarke, S. (1989) *J. Biol. Chem.* **264**, 20131–20139.
8. Haydock, S. F., Dowson, J. A., Dhillon, N., Roberts, G. A., Cortes, J., and Leadlay, P. F. (1991) *Mol. Gen. Genet.* **230**, 120–128.
9. Wu, G., Williams, H. D., Zamanian, M., Gibson, F., and Poole, R. K. (1992) *J. Gen. Microbiol.* **138**, 2101–2112..
10. Wang, A. Y., Grogan, D. W., and Cronan, J. E., Jr. (1992) *Biochemistry* **31**, 11020–11028.
11. Gomi, T., Tanihara, K., Date, T., and Fujioka, M. (1992) *Int. J. Biochem.* **24**, 1639–1649.
12. Fujioka, M. (1992) *Int. J. Biochem.* **24**, 1917–1924.
13. Higgins, D. G., and Sharp, P. M. (1989) *Comput. Appl. Biosci.* **5**, 151–153.
14. Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1978) *in* Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3. National Biomedical Research Foundation, Silver Spring, MD.
15. Lipman, D. J., and Pearson, W. R. (1985) *Science* **227**, 1435–1441.
16. Posfai, J., Bhagwat, A. S., and Roberts, R. J. (1988) *Gene* **74**, 261–265.
17. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., and Roberts, R. J. (1993) *Cell* **74**, 299–307.
18. Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990) *Trends Biochem. Sci.* **15**, 430–434.
19. Scrutton, N. S., Berry, A., and Perham, R. N. (1990) *Nature* **343**, 38–43.
20. Hara, O., and Hutchinson, C. R. (1992) *J. Bacteriol.* **174**, 5141–5144.
21. Schmitt, D., Pakusch, A. E., and Matern, U. (1991) *J. Biol. Chem.* **266**, 17416–17423.
22. Dougherty, D. A., and Stauffer, D. A. (1990) *Science* **250**, 1558–1560.
23. McCurdy, A., Jimenez, L., Stauffer, D. A., and Dougherty, D. A. (1992) *J. Amer. Chem. Soc.* **114**, 10314–10321.
24. Cohen, J. B., Sharp, S. D., and Liu, W. S. (1991) *J. Biol. Chem.* **266**, 23354–23364.
25. Harel, M., Schalk, I., Ehret-Sabatier, L., Bouet, F., Goeldner, M., Hirth, C., Axelsen, P. H., Silman, I., and Sussman, J. L. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 9031–9035.
26. Takata, Y., and Fujioka, M. (1992) *Biochemistry* **31**, 4369–4374.
27. Rafferty, J. B., Somers, William S., Saint-Girons, I., and Phillips, S. E. V. (1989) *Nature* **341**, 705–710.
28. Clarke, S. (1993) *Curr. Opin. Cell Biol.* **5**, 977–983.

Structure

Editors Wayne A Hendrickson (USA) Carl-Ivar Brändén (France)
Minireviews Editor John Kuriyan (USA)

Vol 5 No 4 15 April 1997

London Office
34–42 Cleveland Street,
London W1P 6LB
Tel: (0)171 580 8377
Fax: (0)171 580 8428
E-mail: structure@cursci.co.uk

Editors
Catherine Wild
Victoria Hedges

Production Editor
Damien Vessey

Editorial Assistant
Stephanie Rogers

Editorial Director
Theodora Bloom

Production
Adrienne Hanratty

Illustrator
Matthew McCutcheon
E-mail: matthewm@cursci.co.uk

Advertising
Deborah Breeds

San Francisco Office
211 Hugo Street,
San Francisco,
CA 94122-2603
Tel: 415 566 4880
Fax: 415 566 4594

US Managing Editor
Rebecca Ward

New York Office
111 Eighth Avenue,
Suite 1503,
New York,
NY 10011,
Tel: 212 643 3303
Fax: 212 645 5988
E-mail: info@biomednet.com

Marketing
Mark Tesoriero
Barbara Sullivan

Structure

Structure (ISSN 0969-2126) is published monthly by

Current Biology Ltd,
34–42 Cleveland Street,
London W1P 6LB, UK.

Each volume consists of twelve issues of approximately 150 pages.

Article reprints are available through *Structure's* reprint service. For information contact Joanna Scott at (0)171 323 0323.

The journal is printed on acid-free paper by Pensord Press in the UK.

Advertising information

Advertising is accepted in *Structure*. Further information regarding advertising in the journal, including rate cards, specifications, etc., can be obtained from:

Deborah Breeds,
Current Biology Ltd,
Middlesex House,
34–42 Cleveland Street,
London W1P 6LB, UK
Tel: (0)171 323 0323
Fax: (0)171 436 9293
E-mail: adscb@cursci.co.uk

Structure is indexed and/or abstracted by Current Awareness in Biological Sciences, Current Contents/Life Sciences, Chemical Abstracts, BIOSIS, Excerpta Medica/EMBASE, and MEDLINE/*Index Medicus*.

Subscription information

Subscription rates (Volume 5, 1997, twelve parts) including airspeed delivery

	North/South America ($US)	Rest of World (£UK)
Personal	$235.00†	£150.00‡
Institutional	$695.00†	£450.00‡
Student*	$120.00†	£80.00‡

*Students must give the name of their institution or school, plus the name of their department chairman to qualify for the student rate, which is available for a maximum of two years.
†Canadian subscribers add GST.
‡EU subscribers may be liable to European sales tax.

Periodicals Class postage paid at Middlesex NJ.

US Postmaster:
Airfreight and mailing in the USA,
c/o Pronto Mailers,
PO Box 177,
Middlesex,
NJ 08846,
USA.

Send changes of address to

Structure,
Current Biology Ltd,
c/o BioMedNet USA,
111 Eighth Avenue,
Suite 1503,
New York,
NY 10011,
USA.

Orders

All print orders should be placed with a bookseller or subscription agency, or sent directly to

Turpin Distribution Services Ltd,
Blackhorse Road,
Letchworth,
Hertfordshire,
SG6 1HN,
UK.

Tel: +44 (0)1462 672555
Fax: +44 (0)1462 480947
E-mail: turpin@rsc.org

Send notices of change of address at least 8 weeks in advance, including both old and new address. Cancellations on renewed subscriptions will not be accepted after the first issue has been shipped.

Online subscriptions can be obtained through the BioMedNet library at http://biomednet.co.uk

Subscription queries

All enquiries should be addressed to

Turpin Distribution Services Ltd,
Blackhorse Road,
Letchworth,
Hertfordshire,
SG6 1HN,
UK.

Tel: +44 (0)1462 672555
Fax: +44 (0)1462 480947
E-mail: turpin@rsc.org

In case of particular difficulty please contact

Peter Newmark,
Managing Director,
Current Biology Ltd,
34–42 Cleveland Street,
London,
W1P 6LB,
UK.

Tel: +44 (0)171 580 8377
Fax: +44 (0)171 580 8428
E-mail: peter@cursci.co.uk

Crystal structure of the chemotaxis receptor methyltransferase CheR suggests a conserved structural motif for binding *S*-adenosylmethionine

Snezana Djordjevic and Ann M Stock*

Background: Flagellated bacteria swim towards favorable chemicals and away from deleterious ones. The sensing of chemoeffector gradients involves chemotaxis receptors, transmembrane proteins that detect stimuli through their periplasmic domains and transduce signals via their cytoplasmic domains to the downstream signaling components. Signaling outputs from chemotaxis receptors are influenced both by the binding of the chemoeffector ligand to the periplasmic domain and by methylation of specific glutamate residues on the cytoplasmic domain of the receptor. Methylation is catalyzed by CheR, an *S*-adenosylmethionine-dependent methyltransferase. CheR forms a tight complex with the receptor by binding a region of the receptors that is distinct from the methylation site. CheR belongs to a broad class of enzymes involved in the methylation of a variety of substrates. Until now, no structure from the class of protein methyltransferases has been characterized.

Results: The structure of the *Salmonella typhimurium* chemotaxis receptor methyltransferase CheR bound to *S*-adenosylhomocysteine, a product and inhibitor of the methylation reaction, has been determined at 2.0 Å resolution. The structure reveals CheR to be a two-domain protein, with a smaller N-terminal helical domain linked through a single polypeptide connection to a larger C-terminal α/β domain. The C-terminal domain has the characteristics of a nucleotide-binding fold, with an insertion of a small antiparallel β sheet subdomain. The *S*-adenosylhomocysteine-binding site is formed mainly by the large domain, with contributions from residues within the N-terminal domain and the linker region.

Conclusions: The CheR structure shares some structural similarities with small molecule DNA and RNA methyltransferases, despite a lack of sequence similarity among them. In particular, there is significant structural preservation of the *S*-adenosylmethionine-binding clefts; the specific length and conformation of a loop in the α/β domain seems to be required for *S*-adenosylmethionine binding within these enzymes. Unique structural features of CheR, such as the β subdomain, are probably necessary for CheR's specific interaction with its substrates, the bacterial chemotaxis receptors.

Addresses: Howard Hughes Medical Institute, Center for Advanced Biotechnology and Medicine, Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854, USA.

*Corresponding author.
E-mail: stock@mbcl.rutgers.edu

Key words: bacterial chemotaxis, carboxyl methylation, crystal structure, methyltransferase, receptor modification, *S*-adenosylmethionine

Received: **3 February 1997**
Revisions requested: **18 February 1997**
Revisions received: **26 February 1997**
Accepted: **3 March 1997**

Electronic identifier: **0969-2126-005-00545**

Structure 15 April 1997, 5:545–558

Introduction

The methyltransferases are a large and diverse class of enzymes that catalyze the transfer of methyl groups from *S*-adenosylmethionine (AdoMet) to a wide range of substrates, including small molecules, nucleic acids and proteins. Despite the use of a common cofactor, AdoMet, the mechanism of methyl transfer is not conserved among methyltransferases. This may be reflected in the lack of distinguishing sequence motifs that define the active sites of methyltransferases, comparable to the consensus sequences characteristic of the nucleotide-binding sites of kinases and dehydrogenases. And although AdoMet is the second most commonly used cofactor after ATP, there are relatively few structural descriptions of AdoMet-binding sites.

The three-dimensional structures of several DNA methyltransferases [1–3], an RNA methyltransferase [4] and two small molecule methyltransferases [5,6] have been determined, but structural information regarding protein methyltransferases is lacking. Protein methyltransferases are diverse both with respect to the target amino acid modified by methylation and the proposed role of the modification [7]. Some protein methylations are irreversible and are assumed to have a structural role, such as the methylation of α-amino groups of a variety of N-terminal amino acids or the *N*-methylations of histidine, arginine and lysine sidechains. Methylation at protein carboxyl groups, however, is reversible and appears to function more dynamically. Methylation of aspartyl sidechains

is proposed to be involved in repair of damaged proteins [8,9], and methylations of the C-terminal α-carboxyl groups of eukaryotic proteins [10,11] and of the glutamyl sidechain carboxyl groups in bacterial receptors [12,13] are involved in signal transduction.

Bacterial chemotaxis transmembrane receptors are reversibly modified by methylation of four to five glutamate residues within their cytoplasmic domains (reviewed in [14,15]). The methylation–demethylation reactions are catalyzed by CheR, an AdoMet-dependent protein methyltransferase, and CheB, a methylesterase/amidase. Methylation of the chemoreceptors counterbalances the effects of ligand binding, and contributes to the phenomenon of adaptation by resetting the signaling activity of the receptors despite the continued presence of stimulus. Methylation of the receptors is highly regulated by multiple mechanisms. The regulation of enzyme activity occurs primarily through control of the methylesterase CheB, a response regulator protein that is activated by a two-component phosphotransfer pathway [16,17]. Additionally, both methylating and demethylating reactions are influenced by the specific conformation of the receptors which presumably affects accessibility of the glutamate residues [18–20]. Each specific glutamate residue is methylated at a characteristic rate, which correlates with the magnitude of the effect on chemotaxis of mutation of the specific residue [21,22]. The methyltransferase CheR binds to a five amino acid tail at the C termini of some chemoreceptors; this tail is distant in primary sequence from CheR's sites of methylation [23]. Data suggests that from this tethered position, the methyltransferase methylates other chemoreceptor dimers through inter-dimer interactions.

The structure of the *Salmonella typhimurium* chemotaxis receptor methyltransferase CheR bound to *S*-adenosylhomocysteine (AdoHcy), a product and inhibitor of the methylation reaction, has been determined at 2.0Å resolution. This is the first report of the structure of a protein methyltransferase. Unlike catechol-*O*-methyltransferase and cytosine-DNA methyltransferases, the active site of CheR involves neither metal ions nor a methylcysteinyl intermediate in catalysis. In this respect, CheR is similar to the adenine-DNA methyltransferase and RNA methyltransferase vaccinia protein VP39. Although there are specific differences in topologies, presumably due to the nature of substrates for each enzyme, the structure of CheR confirms that methyltransferases from all four substrate groups (proteins, DNA, RNA and small molecules) share some structural features. Structural analysis of the four classes of enzymes indicates that the AdoMet-binding site is characterized by the specific conformation of a β1/αA loop within the α/β domain, with additional interactions contributed by residues of a linker to an additional domain that is commonly involved in substrate recognition. These features are specific to methyltransferases and distinct from cofactor-binding sites of other nucleotide-binding proteins. In addition to the interest in CheR as an AdoMet-dependent protein methyltransferase, its structure provides a foundation for beginning to explore the complex interactions between a chemotaxis receptor modification enzyme and its multiple substrates.

Results and discussion

Structure determination

The structure of *S. typhimurium* CheR in a complex with AdoHcy was determined at 2.0Å resolution by multiple isomorphous replacement (MIR). The crystals belong to the monoclinic space group $P2_1$. Diffraction data for native and several heavy-atom derivatized crystals were obtained as summarized in Table 1. A native data set for a crystal equilibrated at pH7.0 was also collected. These data merged well with data from the original native crystal grown at pH5.6 (R_{merge}=4.9%), thus a more suitable neutral pH was chosen for preparation of the derivatives. An electron-density map that was calculated with density modified MIR phases is shown in Figure 1.

The atomic model of CheR was refined by using a combination of X-PLOR and ARP procedures (see Materials and methods for details). The final model has good geometry with only one residue, Ser125, outside of the allowed regions of a Ramachandran plot. Ser125 is located in the core of the molecule and is involved in formation of the cofactor-binding site. The unusual backbone conformation of this residue is perhaps unremarkable, because active-site residues commonly acquire unusual phi and psi angles.

The final model contains 2224 non-hydrogen protein atoms corresponding to residues 11 to 284, 26 non-hydrogen atoms belonging to AdoHcy and 110 solvent molecules. The crystallographic R factor for this model is 19.6% for 18035 reflections between 8.0 and 2.0Å resolution, and the free R factor, calculated with 5% of the data, is 27.8%. The root mean squared (rms) deviations from ideal geometry are 0.014Å for bond lengths and 1.7° for bond angles. A summary of the overall quality of the model is presented in Table 2.

The N-terminal ten residues of CheR are not visible in electron-density maps. An N-terminal sequence analysis of the CheR protein, obtained by dissolving the native crystals, indicated that approximately 90% of the protein molecules started at residue Thr2 (indicating removal of the N-terminal Met), with a minor portion of the protein molecules starting at Gln15. These data are in accordance with an apparent disorder of the N-terminal end of the polypeptide chain. Additionally, four amino acids at the C terminus of CheR, residues 285–288, are not observed in electron-density maps.

Table 1

Data collection and MIR phasing.

Compound	Concentration (mM)	Soaking time (days)	Res* (Å)	Com.† (%)	R_{merge}‡ (%)	R factor§ (%)	No. of sites	Phasing# power	R_{cullis}¶ (%)
native pH5.6	-		2.0	90	5.0	-	-	-	-
native pH7.0	-	-	2.0	93	4.9	-	-	-	-
C_2H_5C1Hg	1	3	2.7	76	6.0	23.1	2	2.8	51
C_2H_5C1Hg II	1	3	2.7	85	4.3	24.2	2	2.8	51
Baker's	1	3	2.7	87	5.3	23.3	3	1.4	73
Baker's II	1	3	2.7	89	4.8	34.3	2	1.4	81
DMA	1	3	3.0	78	7.6	42.3	3	1.5	72
K_2PtC1_4	0.5	3 hrs	3.0	85	5.7	34.7	3	1.2	83

*Resolution limit of phasing. †Completeness of data set. ‡$R_{merge}=\Sigma\Sigma|I_{h,i}-<I_h>|/\Sigma\Sigma<I_h>$. §R factor$=\Sigma||I_{PH}|-|I_P||/\Sigma|I_P|$. #Rms amplitude of the heavy atom F/residual lack of closure. ¶$R_{cullis}=\Sigma||F_{PH}-F_P|-|F_H||/\Sigma|F_H|$.

Protein architecture

CheR is a two-domain mixed α/β protein. The α-carbon tracing, ribbon diagram and CPK model of CheR are shown in Figure 2. A topology diagram of CheR is shown in Figure 3. The smaller, N-terminal domain of CheR contains residues 11–90. This domain consists of four perpendicularly packed helices (α1–α4) and an extended terminus formed by residues 11–20, which are oriented away from the domain without interactions with the rest of the molecule. In the crystal, residues 11–20 are packed between two crystallographic symmetry-related molecules, forming few hydrogen bonds and exhibiting high temperature factors (~40 Å²). It is very likely that the conformation of this region is influenced by crystal packing. It is conceivable that the N-terminal twenty residues acquire a different conformation, when interacting with the chemotaxis receptors or in the vicinity of the cytoplasmic membrane. Helix α4 of the N-terminal domain is connected to helix α5 of the C-terminal domain through an extended linker sequence (residues 91–98). A relatively small (~440 Å²) interface between the two domains contains eight hydrogen-bond interactions at the outer rim of the interface, with five hydrophobic residues grouped at the center. Association between the small N-terminal domain, the linker region and the large C-terminal domain is apparent in the CPK representation of the model (Fig. 2c).

The C-terminal domain is composed of residues 99–284. The core of this domain consists of a mixed seven-stranded β sheet that is flanked on both sides by α helices. The

Figure 1

Representative portion of an experimental electron-density map. A stereo image of a region of the final CheR model (black line) is shown superimposed with an MIR/DM electron-density map calculated to 2.7 Å and contoured at 1σ. The figure shows the electron density associated with residues 129–153.

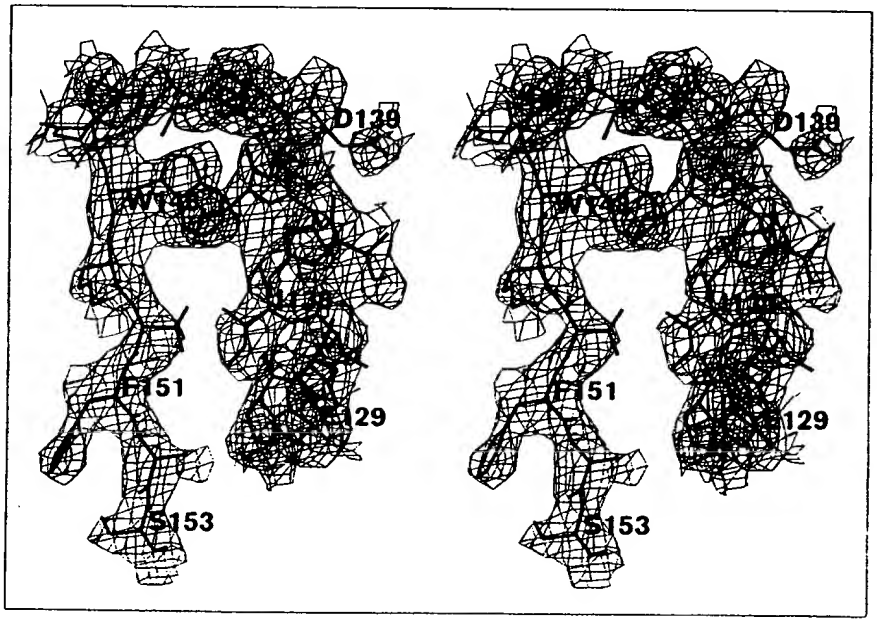


Table 2

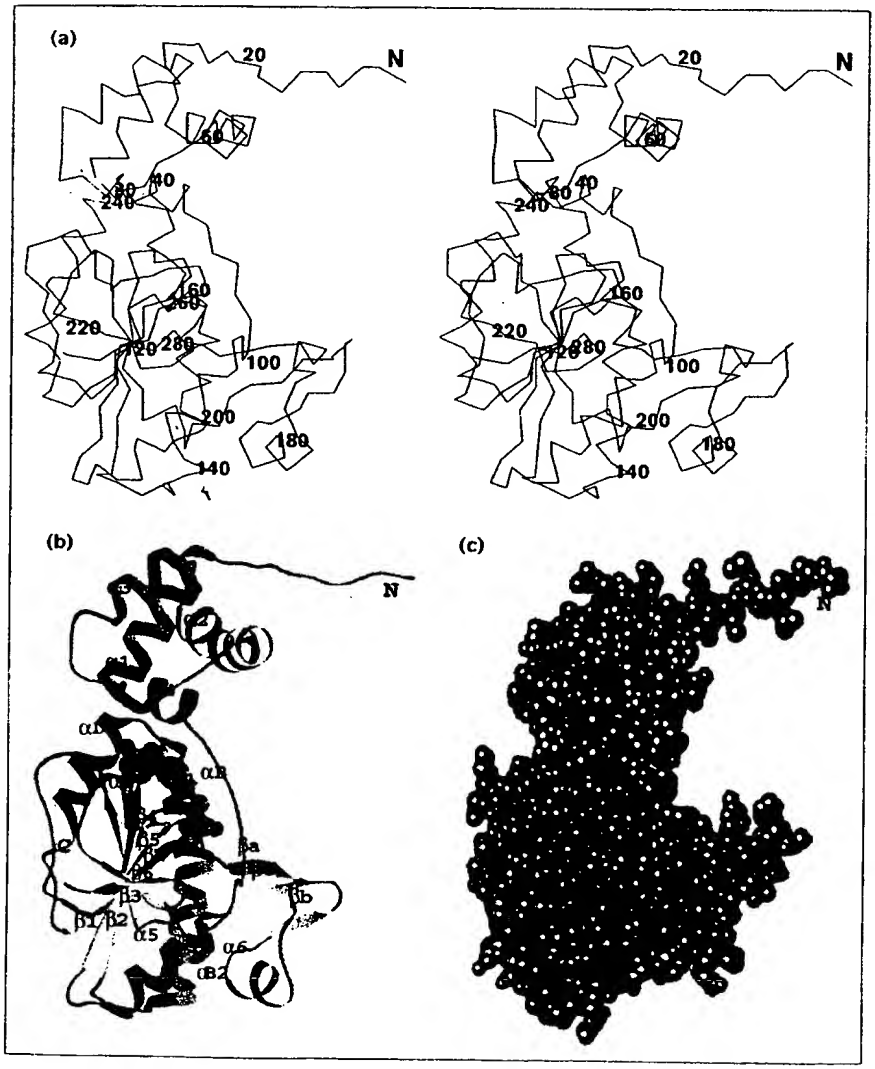
Refinement statistics.

Number of atoms	2360
Resolution (Å)	2.0
Number of reflections	18035
Number of water molecules	110
R factor (%)	19.6
R_{free} (%)*	27.8
Rms bond length (Å)	0.014
Rms bond angle (°)	1.7
Average thermal factors	
Mainchain (Å²)	21.0
Sidechain (Å²)	25.1

*R_{free} was calculated from 5% of reflection data.

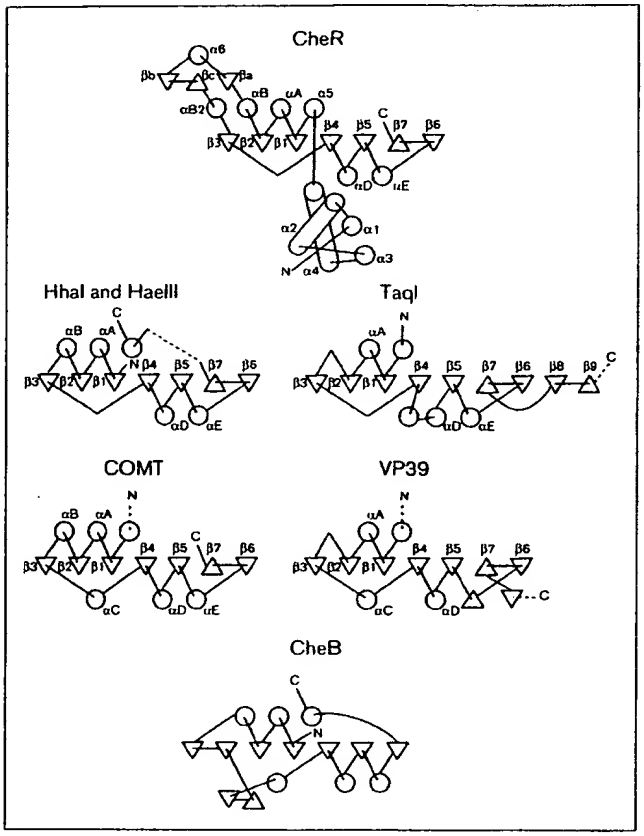
overall topology of the C-terminal α/β domain is common to other previously characterized methyltransferases (Fig. 3). For consistency, in the topological comparison, β strands and α helices are labeled according to the nomenclature previously used for the description of methyltransferases [24,25]. The basic topology represents a variation of a Rossmann-type α/β fold. This was also confirmed by the results of a structural similarity search carried out using DALI [26], which in addition to methyltransferases, identified forty other structures with similarity greater than $Z=3.1\sigma$ (Z is the standard deviations above the mean). These molecules were mostly NAD- and FAD-dependent dehydrogenases and reductases. The methylesterase domain of CheB [27] also shares a structural similarity to CheR ($Z=3.3\sigma$), even though this enzyme does not utilize a nucleotide cofactor. The structural similarity between CheR and CheB might reflect a common evolutionary origin or a requirement for a common scaffold that enables them to react with the same substrates, the chemotaxis receptors.

Figure 2



The three-dimensional fold of CheR. **(a)** A stereo image of the Cα chain of CheR with residue numbers indicated. **(b)** A ribbon diagram (RIBBONS; [56]) of CheR, showing α helices in green, β strands in yellow and 3_{10} helical turns in blue. The AdoHcy molecule bound to CheR is shown in solid spheres. Colors are gray for carbon, blue for nitrogen, red for oxygen and yellow for sulfur atoms. **(c)** A CPK model of CheR, including all atoms. Green spheres indicate residues of the α/β domain and gold spheres indicate residues of the N-terminal helical domain and the linker region. Atoms of the AdoHcy molecule are colored the same as in (b). From this orientation, only gray carbon atoms of AdoHcy are visible. In these figures, the orientations of the model are approximately the same. Figures (a) and (c) were prepared using TURBO.

Figure 3



Schematic topology diagram of CheR, catechol-O-methyltransferase (COMT), *Hae*III, *Hha*I and *Taq*I DNA methyltransferases, RNA methyltransferase VP39 and the catalytic domain of CheB. Helices and strands that are approximately perpendicular to the plane of the figure are represented by circles and triangles, respectively, and the helices that are close to being parallel to the plane of the figure are shown as rods. Dashed lines denote the positions of additional domains. Helical turns that are less than four residues long are not shown as helices in the figure.

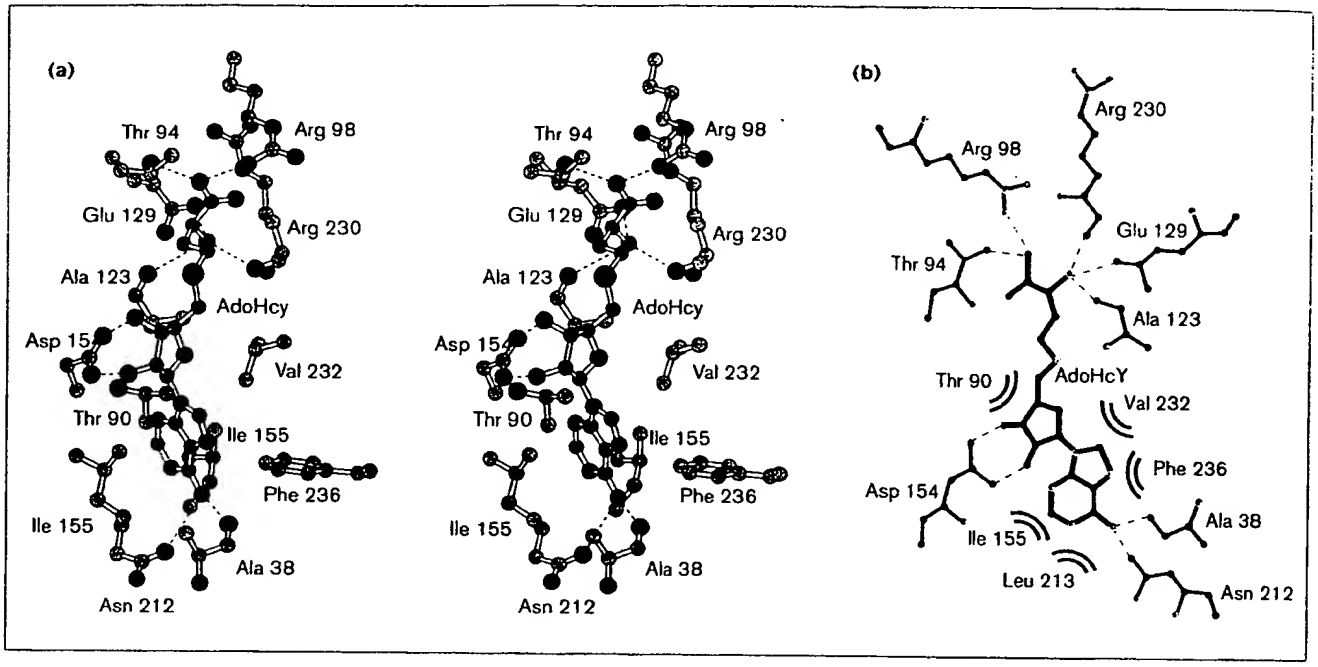
A novel feature of CheR's C-terminal domain is the presence of a small subdomain composed of an α helix and a short antiparallel three-stranded β sheet. This subdomain (residues 166–199) is inserted after αB, at one edge of the β sheet. The subdomain stands almost independently, away from the structure, with only one region involved in interactions with the rest of the α/β domain. Approximately 420 Å^2 of the molecular surface of the α/β domain is buried in the subdomain. At the interface, six hydrophobic residues from the subdomain contribute to the hydrophobic interactions; two of these are tyrosines that also form two out of a total of four hydrogen bonds.

The binding mode of the adenine portion of AdoMet resembles that of NAD in NAD-dependent enzymes, such as alcohol dehydrogenase [29]. The adenine ring is positioned within a hydrophobic pocket sandwiched between two relatively large aliphatic residues (Ile155 and Val232 in the case of CheR), while an acidic residue (Asp154) forms a hydrogen bond with the hydroxyl groups of the ribose ring. In addition, an amino group from the adenine ring forms a hydrogen bond with the backbone carbonyl oxygen of Ala38, within the N-terminal helical domain, and also with the sidechain carbonyl oxygen of Asn212. The homocysteine portion of AdoHcy extends away from the ribose, fitting into an elongated groove in the CheR molecule formed by residues from both the C-terminal domain and the linker region. The oxygen atoms from the carboxyl and amino groups of the homocysteine form hydrogen bonds either with sidechain atoms pointing towards the binding cleft or with the surrounding backbone atoms of CheR. All of the residues that are engaged in hydrogen bonds to the homocysteine are completely conserved among the CheR proteins from a wide variety of bacterial species. The residues forming the hydrophobic environment around the adenine ring, although not completely conserved, are highly similar in all of the CheR sequences.

Because of the involvement of cysteine residues in the catalysis of methyl transfer by cytosine-DNA methyltransferases, the two cysteine residues in *S. typhimurium* CheR, Cys31 and Cys229, have been the focus of a number of mutagenesis and biochemical studies [30]. Substitution of Cys31 with serine resulted in an 80% decrease in methyltransferase activity. Furthermore, it was shown that inactivation of wild-type or Cys229→Ser-mutant CheR by sulfhydryl reagents could be prevented by preincubation of the enzymes with AdoMet. In another report, Cys31 was photolabeled with *S*-adenosyl-L-[methyl-^{3}H]methionine, suggesting that Cys31 was located at, or near, the AdoMet-binding site [31].

Surprisingly, neither Cys31 nor Cys229 are part of the active site. In the three-dimensional structure of CheR, Cys31 resides within the N-terminal domain, ~15 Å away from the AdoHcy-binding site. Thus, it appears that the observed decrease in activity of the Cys31→Ser mutant is a consequence of some effect other than direct participation of this residue in the catalytic reaction. We were unable to obtain crystals of CheR in the absence of cofactor, which may indicate a somewhat flexible nature of the molecule, specifically with respect to the inter-domain connection. In fact, due to the buried nature of the AdoHcy-binding site (Fig. 2c), significant movement of the domains must occur to provide the cofactor access to the binding cleft. Given that the N-terminal domain and linker region also contribute to AdoHcy binding, binding of cofactor may lock Cys31 into a less solvent accessible position and thus protect it from sulfhydryl reagents. In reverse, the Cys31Ser mutation may disrupt inter-domain interactions and thus affect AdoMet binding.

Figure 4



S-adenosylhomocysteine-binding site in CheR. **(a)** A stereo diagram (MOLSCRIPT; [57]) of the AdoHcy-binding site. Only sidechain atoms are included in the figure except for the residues that form hydrogen bonds with AdoHcy through mainchain atoms. Hydrogen bonds are represented by dashed lines. **(b)** A schematic view of the contacts identified in the crystal structure of the CheR–AdoHcy complex. Hydrogen bonds are drawn with dashed lines and covalent bonds are shown as solid lines connecting the solid spheres that denote atoms. Residues within the hydrophobic pocket that accommodates the adenine portion of AdoHcy are represented by parallel curved lines. Sulfur atoms are shown in green, other atom colors are the same as Figure 2.

Comparison of cofactor-binding sites in CheR and other methyltransferases

Overall sequence similarity among methyltransferases that methylate different types of substrates is very weak. Analysis of amino acid sequences of a broad range of AdoMet-dependent methyltransferases revealed only three short regions with sequence similarity [32]. It was suggested that the amino acids in these regions might have a common function, such as the binding of AdoMet/ AdoHcy, or alternatively might be indicators of a common evolutionary origin. More recently, a structure-guided analysis revealed nine conserved sequence motifs specific to DNA amino-methyltransferases [24]. Even though the sequential order of these motifs varies in DNA amino-methyltransferases, their sequence similarity is significant. On the basis of structural and sequence comparisons, it has been proposed that many, and possibly all, AdoMet-dependent methyltransferases have a common catalytic-domain structure [33,25]. Recent publications have examined the primary sequence composition of the AdoMet/AdoHcy-binding site [4] and have addressed its relationship with the NADbinding site of alcohol dehydrogenase [34]. Crystal structures are now available for members of all four major classes of methyltransferases: three DNA methyltransferases — *Hha*I [35], *Taq*I [2] and *Hae*III [3]; an RNA methyltransferase — VP39[4]; two small molecule methyltransferases — catechol-*O*-methyltransferase [5] and glycine methyltransferase [6]; and a protein methyltransferase — CheR. In addition, many more amino acid sequences for a variety of methyltransferases are also available, which enable us to address the questions of cofactor binding, evolutionary origin and possible identification of characteristic primary or tertiary structure elements.

We have superimposed all available methyltransferase structures and have examined the AdoMet/AdoHcy-binding sites in order to carry out a comprehensive and detailed analysis. Table 3 summarizes the structural comparisons of the cofactor-binding sites by listing the residues within a 4 Å distance to the cofactor molecule. We included only the DNA-bound structure of *Hha*I cytosine methyltransferase and not the DNA-free form. In the latter structure, the AdoMet molecule exhibited an inverted orientation within the binding cleft [1], which the authors subsequently suggested was probably a non-physiological phenomenon related to the crystal packing [35]. Additionally, glycine methyltransferase was excluded from our analysis even though it exhibits a great deal of structural

Table 3

Methyltransferase residues involved in interactions with AdoMet/AdoHcy.

Binding-site feature	CheR	COMT*	*Hha*I†	*Hae*III†	*Taq*I§	VP39#
Hydrogen bonds with AdoHcy atoms¶						
N6;Ade	O;Asn212	O;Gln120	O;Asp60	O;Asp50	O;Asp89	O;Val116,m
N6;Ade	O;Ala38,m	O;Ser119	–	–	–	–
N;HomoCys	O;Ala123,m	O;Gly66,m	–	–	–	O;Gly68,m
N;HomoCys	O;Glu129	O;Asp141	–	–	O;Glu45	O;Asp138
N;HomoCys	O;Arg230,m	O;Ser72	–	–	–	O;Tyr66
O1;HomoCys	–	–	N;Gly23,m	N:Gly12,m	–	–
O1;HomoCys	–	–	O;Ser305	–	O;Glu22	–
O2;HomoCys	O;Thr94	O;Ser72	O;Ser305	–	O;Thr23	N;Gln39
O2;HomoCys	N;Arg98	N;Val42,m	N;Leu21,m	N;Ala10,m	–	N;Gly72,m
O2;Ribose	O1;Asp154	O1;Glu90	O1;Asp40	O1;Glu29	O1;Glu71	O1;Asp95
O3;Ribose	O2;Asp154	O2;Glu90	O1;Asp40	O1;Glu29	O2;Glu71	O2;Asp95
Hydrophobic interactions with adenine ring						
	Ile155	Met91	Phe18	Phe7	Ile72	Ile67
	Leu213	His142	Trp41	Tyr30	Phe90	Phe115
	Val232	Trp143	Ile61	Ile51	Pro107	Val116
	Phe236		Pro80	Pro70	Phe146	Val139
Residues in β1/αA loop††	**AAASTGE**	**LGAYCGY**	**LFAGLGG**	**LFSGAGG**	**PACAHGP**	**IGSAPGT**

*Catechol-O-methyltransferase. †*Hha*I cytosine methyltransferase. ‡*Hae*III cytosine methyltransferase. §*Taq*I adenine methyltransferase. #VP39 vaccinia protein RNA methyltransferase. ¶Format: methyltransferase atom; residue with m indicating a mainchain atom. ††Bold residues belong to the loop, the first and the last residues flanking the loop correspond to the end of β1 and the beginning of αA, respectively; this region corresponds to the FGLGG sequence in alcohol dehydrogenase.

similarity with other methyltransferases. AdoMet binds very differently in this enzyme, primarily due to the multimeric nature of glycine methyltransferase and the presence of an additional domain [6]. Glycine methyltransferase also binds tetrahydrofolate and polycyclic aromatic hydrocarbon molecules, indicating a much less AdoMet-specific binding character. The active site of this enzyme presumably represents an entirely different class of cofactor-binding sites.

The only residues common to all AdoMet-binding sites are an aspartate or a glutamate, which forms hydrogen bonds to the hydroxyl groups of the ribose and, with the exception of VP39, an aspartate, glutamate or asparagine, which forms hydrogen bonds to the amino group of the adenine. These residues, however, are also found in alcohol dehydrogenase and other enzymes that bind NAD, a cofactor with adenine and ribose moieties identical to AdoMet/ AdoHcy. Interestingly, in RNA methyltransferase VP39, there is also an aspartate residue in a corresponding position. This aspartate, however, points away from the AdoMet, and instead the adenine amino group forms a hydrogen bond with the carbonyl oxygen of the neighboring residue. Similarly to the adenine of NAD in NAD-dependent enzymes, the adenine ring of AdoMet/AdoHcy in all methyltransferases is situated within a hydrophobic pocket of variable residues. There is also variability in the methyltransferase residues that participate in hydrogen bonds with the carboxyl and amino groups of the methionine portion of AdoMet/AdoHcy.

Based on examination of the structures, it appears that the size and shape of the AdoMet-binding clefts are very similar in all methyltransferases, despite the apparent lack of sequence identity. Importantly, in all of the methyltransferases, the binding cleft is formed mainly by residues coming from the α/β domain, with some additional contributions from residues within the linker region that connects the α/β domain to an additional domain, which is commonly involved in determining substrate specificity. The loop connecting the first β strand in the α/β domain to the following a helix is well characterized in NAD-binding enzymes as it contains the conserved sequence motif Gly–X–Gly–Gly, in which the last glycine is the first residue of the α helix. In methyltransferases, the sequence of this loop is much less conserved. The loop is two residues longer compared to that of alcohol dehydrogenase, such that it greatly affects the position of the connecting α helix, and it directs the shape of the bed of the AdoMet-binding cleft over which the methionine portion of the cofactor is positioned. The sequences involved in forming this loop are listed in Table 3. A common feature of all methyltransferases is that the loop ends with a glycine residue and in all but RNA methyltransferase VP39, the third residue within the loop displays specific phi/psi values (approximately phi=50 and psi=−130) falling into a disallowed region of the Ramachandran plot, regardless of the amino acid type. The unusual phi/psi values are the consequence of these residues being part of a hairpin type II structure. In VP39, the fourth residue in

the loop is *cis* proline; this changes the hydrogen-bonding scheme of the loop and introduces a small displacement in the αA helix, not observed in the other methyltransferases. However, the overall shape of the β1/αA loop in VP39 is still highly similar to all other methyltransferase structures.

Despite the highly similar topologies of methyltransferases and NAD-binding enzymes, the β1/αA loops, which are of great importance for both classes of enzymes, exhibit very different conformations. We have examined this region in all available structures containing nucleotides in Rossmann-type α/β folds. The majority of structures contain a β1/αA loop that is three residues long, similar to alcohol dehydrogenase; a smaller number of structures contain a loop that is four residues long, similar to that found in cholesterol dehydrogenases. Within each of the groups represented by alcohol dehydrogenase and cholesterol dehydrogenases, the backbone atoms of the 16 residues that form the β1-loop-αA region overlay very closely with rms deviations ranging from 0.33 Å to 0.86 Å and 0.67 Å to 1.03 Å, respectively. In the structure of enoyl acyl carrier protein reductase, although the loop is six residues long, the β strand and α helix overlay well with the other enzymes (rms deviation 0.67 Å). Structural alignment of the β1-loop-αA region of CheR (residues 118–134) with corresponding residues of the other methyltransferase structures gives backbone rms deviation values from 0.69 Å for CheR and catechol methyltransferase to 1.06 Å for CheR and *Taq*I methyltransferase. Figure 5 shows the α-carbon models of these loops from the methyltransferases and the loop region of the alcohol dehydrogenase structure, aligned by superimposing the β1 strand and the first two loop residues of the corresponding structure of CheR. The larger loop found in the methyltransferases creates a proper space for binding of AdoMet and, within the cleft, the α-amino and α-carboxyl groups of methionine form hydrogen bonds with whatever protein atoms are available within the appropriate distances. The loop itself, in different methyltransferases, not only dictates the shape of the binding cleft, but actually provides the specific backbone atoms involved in hydrogen-bond formation with the cofactor or it specifically positions the sidechains at the end of β1 or at the beginning of αA for formation of hydrogen bonds (Table 3). Even though there is sequence conservation within the DNA methyltransferases, the comparison of diverse and unrelated methyltransferases reveals that it is not exclusively the specific sequence, but rather the specific length and conformation of the β1/αA loop that allows for AdoMet to bind. In most of the methyltransferases, the presence of a second domain also contributes to formation of the binding cleft and the overall binding energy of AdoMet.

It is now possible to examine the previously identified regions of primary sequence similarity [32], with respect to the three-dimensional structures of the methyltransferases. Two sequences, designated consensus regions II and III, are located at the beginnings of β strands four and five, respectively, and are distant from the AdoMet-binding site. Examinations of a larger number of methyltransferase sequences now suggests that these regions are much less conserved than indicated by the original alignment. The residues that are fairly conserved (aspartate from region II and lysine/arginine from loop region III) sometimes, but not always, form a salt bridge. The aspartate residue is also present in a large number of dehydrogenases. Another sequence, designated consensus region I, was identified primarily among DNA methyltransferases. The residues in region I are involved in formation of the specific β1/αA loop, discussed above. As we have already concluded, even in methyltransferases in which the specific consensus region I sequence is not present, the length and conformation of the loop are conserved. It is most likely that the consensus region I sequence reflects a common origin of these enzymes rather than a necessary requirement for AdoMet binding.

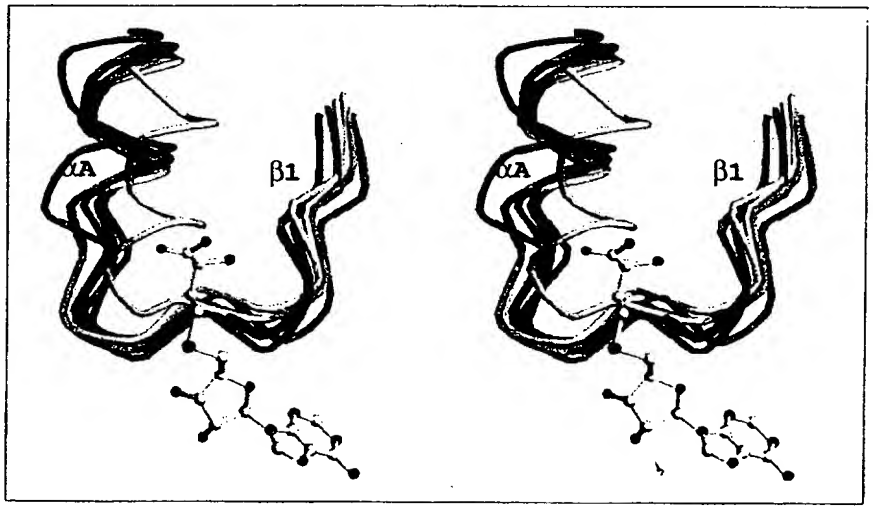
Comparison with CheR homologs from other organisms

The nucleotide sequences of a number of methyltransferase genes displaying sequence similarity with CheR have been determined from a diverse array of bacterial species. An alignment of the predicted amino acid sequences together with the secondary structure of *S. typhimurium* CheR is shown in Figure 6. Overall the sequences exhibit significant similarity, with identity to *S. typhimurium* CheR ranging from 87% for the closely related CheR methyltransferase from *Escherichia coli* to 25% for a hypothetical protein from *Campylobacter jejuni*. Several features are apparent from the alignment. Sequence similarity among the large C-terminal domains is much stronger than in the small N-terminal domain. The region comprising the N-terminal domain differs in length in proteins from different organisms. In some proteins, the sequence extends beyond the C terminus of *S. typhimurium* CheR. As might be expected, there are variations in the lengths of regions corresponding to a few of the surface loops. Notably, there is weak similarity in the region of sequence corresponding to the antiparallel β-sheet subdomain of *S. typhimurium* CheR (residues 167–200). However, as was discussed previously in relation to AdoHcy binding, the residues that are involved in forming hydrogen bonds to AdoMet/AdoHcy are strictly conserved, and residues that form the hydrophobic pocket for the adenine moiety of the cofactor have conserved aliphatic sidechains.

The biological and structural implications of the sequence comparison is that the overall fold, domain structure and mode of cofactor binding are conserved among these methyltransferases. With the exception of the proteins from *Rhodobacter capsulatus*, *C. jejuni* and *Pseudomonas fluorescens*, the proteins presented in Figure 6 have been shown to be associated with chemotaxis or other motility systems, and are presumably involved in modification of receptors or receptor homologs. The substantial divergence in some

Figure 5

Comparison of the cofactor-binding loops of the methyltransferases and alcohol dehydrogenase. A stereo view of α-carbon traces of the β1/αA cofactor-binding loops from six different methyltransferase crystal structures (CheR residues 119–134, yellow; catechol-O-methyltransferase, peach; *Hha*I DNA methyltransferase, dark blue; *Hae*III DNA methyltransferase, magenta; *Taq*I DNA methyltransferase, green; and RNA methyltransferase VP39, gray), and alcohol dehydrogenase (light blue) are superimposed. The structure of the AdoHcy molecule from the complex with CheR is also shown as a ball-and-stick model, with the colors: yellow for carbon, blue for nitrogen, red for oxygen and green for sulphur. The figure was prepared using RIBBONS.



regions of the methyltransferase sequences may allow for recognition of specific receptors. In different organisms, chemotaxis receptors exhibit significant variability thus the specific sequences of the methylation enzymes might allow them to adopt conformations necessary to recognize unique features of their substrates. The lower levels of sequence conservation in the N-terminal domain and in the antiparallel β-sheet subdomain of the C-terminal domain are consistent with the hypothesis that these regions may be involved in interactions with the chemoreceptors.

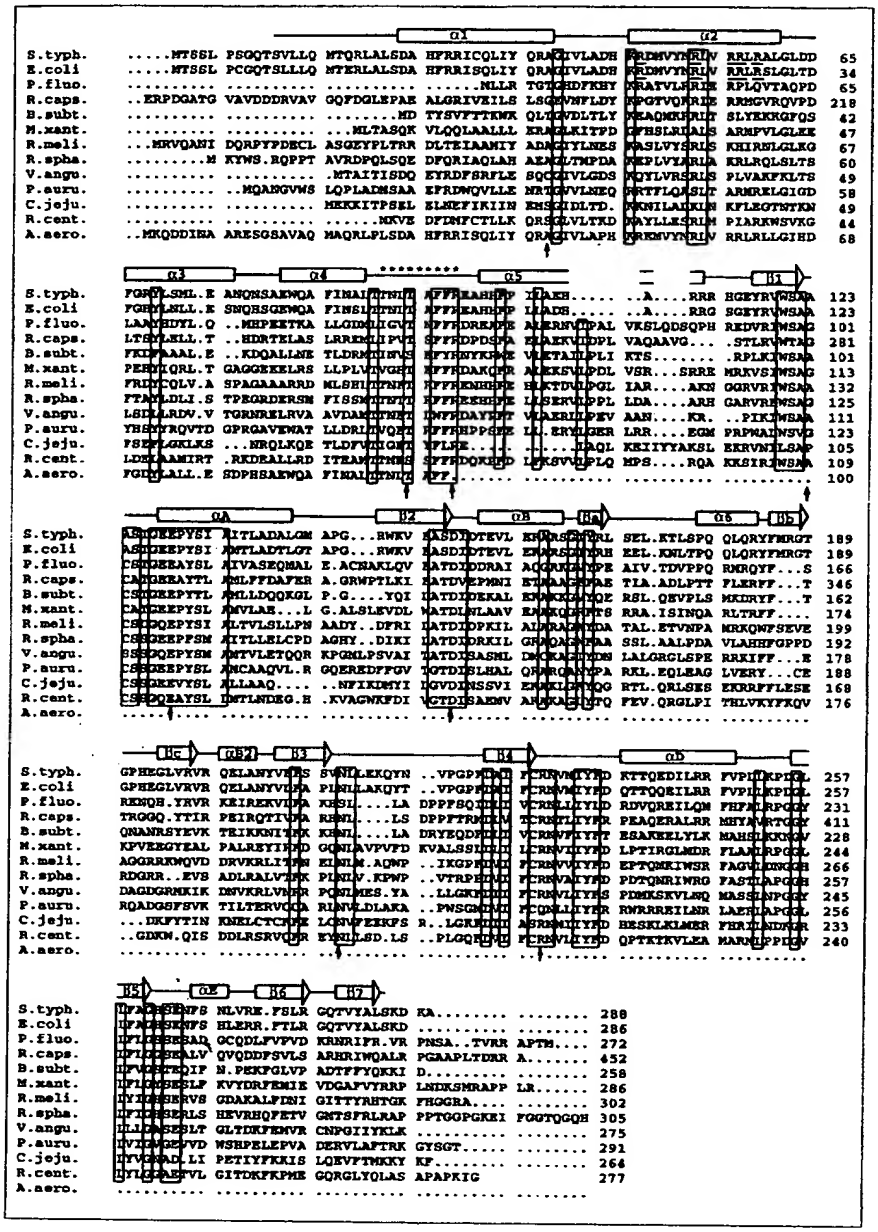
Interaction between the methyltransferase and its receptor substrate

CheR catalyzes the carboxyl methylation of specific glutamyl sidechains within the cytoplasmic domains of the chemotaxis transmembrane receptors of enteric bacteria, such as *E. coli* and *S. typhimurium*. The substrates of CheR are somewhat diverse — it recognizes four to five different glutamyl residues within each of at least five different receptors. Comparison of these sites has revealed a methylation consensus sequence, Glu–Glu–X–X–Ala–Ser/Thr, with methylation occurring at the second glutamyl residue [36–39]. Although there is a significant amount of sequence similarity, and presumably structural homology, at the different methylation sites [40,38,41], there must also be differences as reflected by the different rates of methylation observed at each site [21,22,42].

In the methyltransferase–substrate (inhibitor) complexes, for which structures have been determined, the smaller domain of the methyltransferase is involved in binding the substrate and positioning it appropriately for presentation of the substrate methylation site to the methyltransferase active site located in the large α/β domain. In the *Hha*I and *Hae*III methyltransferases, DNA binds in a cleft formed between the small and large domains [35,3], whereas the catechol *O*-methyltransferase binds the relatively small inhibitor 3,5-dinitrocatechol in a shallow groove of the large domain with hydrophobic interactions contributed by a residue from a loop of the small domain [5]. Despite the common participation in binding substrate, the folds of the small domains of these enzymes are quite distinct. By analogy with these methyltransferases, the N-terminal domain of CheR may participate in interactions with regions of the receptors that contain the methylation sites.

The methyltransferase CheR binds tightly to the cytoplasmic domains of the chemotaxis receptors [43,44]. The binding site has recently been localized to a five amino acid sequence, Asn–Trp–Glu–Thr–Phe, located at the extreme C termini of the *E. coli* and *S. typhimurium* aspartate receptors (Tar) and the *E. coli* serine receptor (Tsr). The intact receptor and a synthetic pentapeptide of the binding site motif exhibit similar binding affinities ($K_a \approx 4 \times 10^5 M^{-1}$) [23]. On the basis of this observation and the lack of conservation of this binding motif within all of the methylated receptors, a model for intermolecular receptor methylation has been proposed. In this model, CheR is tethered to the C-terminal tail of one receptor dimer, from where it methylates an adjacent, and perhaps heterologous, receptor dimer in the membrane. Although at this time we have no knowledge of the receptor peptide binding site in the methyltransferase, the antiparallel β sheet insertion in the large α/β domain provides an intriguing candidate. This subdomain, appears to be less ordered than the rest of the molecule, with average backbone atoms B values of 28 Å^2 (residues 170–195) as compared to 20 Å^2 for the rest of the α/β domain. The β subdomain does not appear to be an integral part of the overall fold, and has only minimal interactions with the

Figure 6



Sequence alignment of CheR homologs from different organisms. The secondary structural elements of *S. typhimurium* CheR are shown as arrows (β strands), bars (α helices) and lines (connecting loops) aligned above the amino acid sequences. Sites where identical amino acid residues occur in 75% or more of the proteins are boxed. Residues forming the linker are labeled with stars. Arginine residues from helix α2, which are postulated to interact with the chemotaxis receptors in *E. coli* and *S. typhimurium*, are underlined. Vertical arrows indicate residues that form hydrogen bonds with AdoHcy. Accession numbers for the sequences are: *Salmonella typhimurium* (SP, P07801); *Escherichia coli* (SP, P07364); *Pseudomonas fluorescens* (GB, L29642); *Rhodobacter capsulatus* (SP, Q02998); *Bacillus subtilis* (GB, M80245); *Myxococcus xanthus* (SP, P31759); *Rhizobium meliloti* (GB, U13166); *Rhodobacter sphaeroides* (PIR, S47261); *Vibrio anguillarum* (GB, U36378); *Pseudomonas aeruginosa* (GB, U11382); *Campylobacter jejuni* (SP, P45676); *Rhodospirillum centenum* (GB, U64519); and *Aerobacter aerogenes* (SP, P21824). SP = SWISSPROT; GB = GenBank. The program PILEUP from the Genetics Computer Group Wisconsin package version 8 was used for the alignments [58].

large domain (see Fig. 2 and description of protein architecture). Nor can the presence of this subdomain be rationalized in terms of catalysis at the active site, as it makes no contacts to the bound cofactor.

We have recently obtained co-crystals of CheR in complex with the N-acetylated receptor pentapeptide (SD, unpublished results). Notably, crystals of the complex cannot be obtained under conditions used to grow crystals of CheR alone. Furthermore, crystals of the complex have a different morphology and cell constants than those of CheR, suggesting perhaps that the peptide influences either the conformation of CheR or specific lattice contacts between methyltransferase molecules. The solution of the structure of the CheR–receptor-peptide complex should elucidate the peptide-binding site on CheR and provide a foundation for further investigations of methyltransferase–receptor interactions.

Catalytic mechanism

Biochemical analyses of the *E. coli* and *S. typhimurium* methylation systems have indicated that the methylation

reaction catalyzed by CheR proceeds by a random mechanism [28]. The turnover number of 10 min^{-1} and the absence of a covalent enzyme–substrate intermediate support a reaction model involving direct transfer of a methyl group from AdoMet to the glutamyl carboxyl of a specific sidechain of a chemotaxis receptor. It is also known that direct methyl transfer reactions occur by an S_N2 type of mechanism [45], thus the reaction rates are strongly dependent on characteristics of the nucleophile, in this case, the receptor glutamate carboxylate oxygen. The active site for methyl transfer necessarily involves residues from both the region of the receptor that is methylated and CheR. On the basis of the knowledge of the reaction mechanism and insights provided by the structure of the CheR-AdoHcy complex, we propose the following: firstly, binding of the methylation region of the chemotaxis receptor occurs within the wide opening of CheR, formed by the central AdoMet-binding site, flanked on either side by the β subdomain and the N-terminal helical domain; secondly, formation of the Michaelis complex follows a conformational change in CheR and possibly in the substrate as well; and thirdly, reaction rates are dependent on the specific presentation of the receptor glutamate residues to the interaction surface of CheR, which probably involves the positively charged helix α2.

In the structure of the CheR–AdoHcy complex, the cofactor is fully buried in the protein-binding cleft with no atoms accessible to solvent. The inclusion of a methyl group at the sulfur atom of homocysteine, which would correspond to a molecule of AdoMet, might require some movement of the surrounding sidechains, such as Tyr235 or even Arg89. Binding of the receptor substrate could entail additional conformational changes, necessary to properly position and orientate the reactive groups. Such movement has previously been observed in the complex of DNA–*Hha*I DNA methyltransferase, in which binding of DNA to the enzyme was associated with a significant conformational change within the DNA-binding domain as well as distortion of the DNA helix, with the methylatable base completely flipping out from the helix [35]. The intramolecular surface of interaction between N-terminal and C-terminal domains of CheR is relatively small (about 400 $Å^2$) and is primarily built of several hydrogen bonds and a few hydrophobic interactions. These domains could potentially move in some type of the hinge motion, forming a complex with the receptor substrate or perhaps opening the active site to allow for binding of AdoMet.

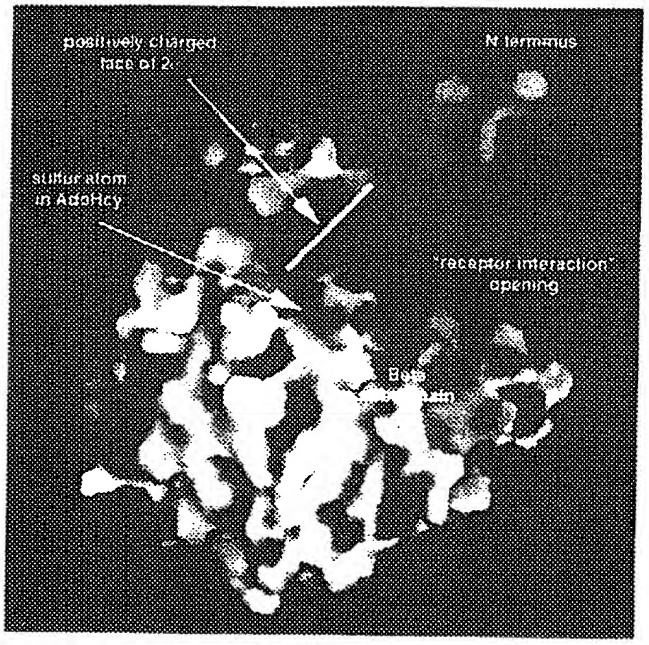
The observed differences in rates of carboxyl methylation reactions at different glutamate residues can be explained by two different effects, or perhaps by a combination of both. One possibility is that the reaction rates are determined by the strengths of the nucleophile — carboxyl oxygen or by the mobility of a leaving methyl group, both of which can differ depending on other factors. This would suggest the involvement of an additional residue that would specifically stabilize a transition state. For example, an oxygen of a glutamate or tyrosine residue might interact with the positively charged sulfur atom, as is postulated in the reaction mechanism of glycine methyltransferase, or some positively charged residue might interact with the glutamyl substrate (analogous to the role of Mg^{2+} in catechol methyltransferase). The presence of a receptor methylation consensus sequence Glu–Glu*–X–X–Ala–Ser/Thr (X = any residue; Glu* = methylated residue) implies, however, that the active-site residues contributed by both CheR and the receptor are similar for all four methylatable glutamate residues. Hence, it is more likely that the reaction rates are determined by another mechanism, such as the availability, specific orientation and positioning of the glutamate residues for methylation, rather than by a different composition of active-site residues. This interpretation is also supported by the mutagenesis studies of Shapiro and Koshland, which have shown that the shorter carboxylate sidechain of aspartates cannot be methylated and that substitution of a methylatable glutamate with aspartate drastically decreases the rate of methylation on the glutamate residue N-terminal to the mutation [21].

Calculated electrostatic potentials (GRASP) [46] of the CheR surface revealed that the proposed surface area of interaction with the receptor is positively charged overall. This region includes an approximately 12 Å wide positively charged surface, leading to the AdoHcy-binding pocket (Fig. 7). Charges in this area are provided by helix α2; three arginine residues (53, 56 and 59) are lined up on one face of the helix and are oriented towards the proposed receptor interaction opening. Methylatable glutamate residues on the receptor are followed by two to three non-methylatable glutamates in the first α helix and preceded by four to five non-methylatable glutamates in the second α helix of the predicted antiparallel coiled coil of the receptor cytoplasmic domain [41]. It is possible that these residues, all of them seven residues apart, form a negatively charged surface that is involved in specific interactions with the positively charged surface of CheR and that this complementary electrostatic interaction serves to position the methylatable glutamates within the active site.

It should be noted that helix α2 is significantly distorted, due to the hydrogen bonds formed by sidechain atoms of its residues and the mainchain atoms of the same or neighboring molecule (lattice contacts). This observation suggests a flexibility of this helix and its potential for adopting slightly different conformations that might be necessary for interaction with the receptor.

According to the model of Wu *et al.* [23], the receptor substrate, via its C-terminal pentapeptide, will bind

Figure 7



Calculated electrostatic potential surface of the CheR–AdoHcy complex. Electrostatic potential on the molecular surface of CheR is color coded as red for negatively charged, blue for positively charged and gray for neutral regions. An arrow points to the position of the sulfur atom in AdoHcy that is not solvent exposed. The figure was generated using GRASP [46].

CheR at an independent site (perhaps involving the CheR β subdomain) and from that tethering point allow for the dynamic interaction of different receptor glutamate residues with the AdoMet bound to CheR. Different positions of the substrate glutamates on the helix of the predicted coiled coil might place particular residues in a more favorable position for methyl transfer than others, even though all contain a common consensus sequence for interaction with CheR. This scenario does not eliminate the potential role for specific CheR or receptor residues in stabilization of a transition state. However, a detailed description of the complete active site and reaction mechanism will only be possible with a combination of structural information for the receptor–CheR complex and further mutagenesis data.

Biological implications

Bacterial chemotaxis receptors are transmembrane proteins that detect environmental stimulants through their periplasmic domains and transduce the signal to proteins downstream of the cytoplasmic two-component signaling system. In addition to the effect of ligand binding, the signaling output of the chemotaxis receptors is modulated by the level of covalent modification — methylation of specific glutamate residues on the cytoplasmic domains. Methylation of the chemotaxis receptors counter-effects ligand binding. Methylation of the receptor glutamate carboxylate groups is catalyzed by CheR, a methyltransferase that uses an *S*-adenosyl-L-methionine (AdoMet).

CheR belongs to a broad class of AdoMet-dependent methylating enzymes that are involved in numerous processes, such as the synthesis of antibiotics, the metabolism of neurotransmitters and the covalent modification of DNA, RNA, and proteins. Previously, structural analyses have been carried out on the small molecule, RNA and DNA methyltransferases. The crystal structure of CheR, determined at 2.0 Å resolution, is the first structure of a protein methyltransferase.

Protein methylation occurs at amino groups, where it is irreversible and presumably has a structural role, or at carboxyl groups where it is reversible and is postulated to be involved in protein repair and signal transduction. CheR is a two-domain protein carboxylmethylating enzyme. The basic structural elements of the polypeptide topology of the larger α/β C-terminal domain of CheR are similar to those of DNA methyltransferases, an RNA methyltransferase, and catechol-*O*-methyltransferase. The unusual topological features of CheR are its helical N-terminal domain and the presence of a small three-stranded antiparallel β subdomain inserted within the C-terminal domain.

The CheR structure allows for a comprehensive comparison of AdoMet-binding sites among DNA, RNA, small molecule and protein methyltransferases. Although there is little conservation of amino acid sequence motifs, the AdoMet-binding sites are structurally conserved. The specific length and conformation of a loop within the α/β domain (the β1/αA loop in CheR), rather than a specific sequence, are implicated in providing the specificities for AdoMet binding by methyltransferases. The unique structural elements of CheR — specifically, the antiparallel β subdomain and N-terminal domain of the methyltransferase — are probably involved in both tethering the methyltransferase to the receptor and in the interaction with the receptor methylation region. The structure of CheR lays the foundation for detailed examination of the interactions between a modification enzyme and its multiple chemotaxis receptor substrates.

Materials and methods

Crystallization

Crystals of *S. typhimurium* CheR were obtained as previously described [47]. The crystals were prepared using the hanging drop method with a reservoir solution of 1.2 M ammonium sulfate, 2% PEG 400, 25 mM sodium citrate, pH 5.6. For the purpose of introducing heavy atoms, native crystals were transferred to a solution containing 1.4 M ammonium sulfate, 2% PEG, 25 mM sodium citrate, pH 7.0. Subsequently, the crystals were transferred to drops that additionally contained appropriate concentrations of heavy metals.

Data collection

The crystals belong to the monoclinic space group $P2_1$, with cell dimensions a=55.0Å, b=48.0Å, c=63.2Å and β=112.3°. There is one CheR molecule in the asymmetric unit and the solvent content is 40%.

All data sets were collected at room temperature on an R-Axis II (Rigaku) image plate detection system mounted on a Rigaku rotating anode generator operated at 50mA and 100kV, with double mirror focusing (Molecular Structure Corporation). Data were integrated using the program Denzo [48]. All data were merged, scaled and truncated with the Rotavata/Agrovata and Truncate programs in the CCP4 suite [49]. Potential heavy-atom derivative data were scaled to the native data set using the program Scaleit in CCP4.

MIR phasing, model building and refinement

Difference Patterson maps, calculated with derivative and native pH5.6 data, were generated and examined using the programs Topdel, Fsfour and Mapview from the PHASES program suite [50]. Coefficients for a cross-difference Fourier synthesis were calculated with the program Mrgdf (PHASES). Heavy-atom sites were identified by examination of difference Patterson maps, and subsequently validated by cross-difference analysis.

We were able to obtain only mercury and K_2PtCl_4 derivatives, despite an extensive search through a variety of compounds. All of the mercury derivatives had two major and mainly overlapping sites, which in the final model could be correlated with the positions of Cys31 and Cys229. Combination of all mercury derivatives and solvent flattening of the resulting phases yielded a poor electron-density map that, apart from detectable features for two α helices, was not interpretable. Platinum atoms from K_2PtCl_4 bound very strongly to CheR crystals with the main site at 0.0, 0.25, 0.45 (relative to the main Hge specific positions of the Pt atoms, Pt-derived phases were pseudo-centrosymmetric and when used in the cross-difference Fourier method they gave an ambiguous solution in the Y direction for the mercury positions, even though the X and Z coordinates of the mercury atoms were confirmed. An interpretable electron-density map was obtained with phases derived by combination of all of the derivatives followed by density modification procedures.

The positions and occupancies of the heavy atoms for each of the derivatives were refined individually in MLPHARE (CCP4) [51] and then combined, giving a figure of merit of 0.71. The handedness of the structure was determined using anomalous data for one of the mercury derivatives. MIR phases were calculated for data between 35.0 and 3.0Å resolution. These phases were subjected to density modification using the program DM (CCP4) [52], which included solvent flattening, histogram matching and Sayer's equation options. Exclusion of any of the mercury derivatives resulted in much poorer quality of the phases.

MIR/DM phases were used to calculate electron-density maps with data from 25–4Å and 25–3Å. A number of α helices and β strands were clearly identified. Both maps were used for the initial Cα-chain tracing, aided by the bones option within the graphics program Turbo-Frodo [53]. Turbo was used for all of the map interpretation, model building and subsequent map fittings. There was several regions of ambiguity within the maps. Great improvement in the quality of the electron-density map was achieved by using the iterative skeletonization process implemented in DM. At this time, the phases were extended to 2.7Å. The resulting electron-density map enabled us to trace 243 out of 287 residues. Residues 1–11 and 285–287 were missing, and the region between residues 166 and 200 fell within disordered electron density that was difficult to interpret. The partial model was taken through several rounds of positional and simulated-annealing refinement in X-PLOR [54], followed by refitting of the electron-density maps. Model rebuilding was done by examining SIGMAA weighted maps, and $3F_o-2F_c$ and original MIR/DM electron-density maps. In the first few cycles, refinement was carried out with data from 8–3.0Å, after which the resolution was extended to 2.4Å. Calculated phases from the improved model revealed better features in a $3F_o-2F_c$ electron-density map, such that it was possible to trace residues 167–199 and confirm sidechain assignments. The new model, which included residues 11–284 as well as AdoHcy, was refined in X-PLOR starting with 2.4Å and then including data to 2.0Å, resulting in a working factor of 0.277 and a Free R factor of 0.367, without any water molecules and B factors included in refinement. This model was then refined against the native data set at pH7.0, which has slightly better statistics and higher completeness compared to the pH5.6 data set. Water molecules were added using unrestrained and restrained refinement of ARP [55]. Water molecules were carefully examined and another round of X-PLOR refinement was carried out yielding the final model.

Accession numbers

Atomic coordinates have been deposited in the Brookhaven Protein Data Bank, with the code 1af7.

Acknowledgements

We thank Donald Winkelmann for helpful comments on the manuscript. This work was supported in part by a grant from the NSF (MCB9258673) and funding from the WM Keck Foundation for support of structural biology computing at the CABM. AMS is an assistant investigator and SD is a research associate of the Howard Hughes Medical Institute.

References

1. Cheng, X., Kumar, S., Posfai, J.W. & Roberts, R.J. (1993). Crystal structure of the *Hha*I DNA methyltransferase complexed with *S*-adenosyl-L-methionine. *Cell* **74**, 299–307.
2. Labahn, J., *et al.*, & Saenger, W. (1994). Three-dimensional structure of the adenine-specific DNA methyltransferase M•*Taq* I in complex with the cofactor *S*-adenosylmethionine. *Proc. Natl. Acad. Sci. USA* **91**, 10957–10961.
3. Reinisch, K.M., Chen, L., Verdine, G.L. & Lipscomb, W.N. (1995). The crystal structure of *Hae*III methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* **82**, 143–153.
4. Hodel, A.E., Gershon, P.D., Xuenong, S. & Quiocho, F.A. (1996). The 1.85Å structure of vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. *Cell* **85**, 247–256.
5. Vidgren, J., Svensson, L.A. & Liljas, A. (1994). Crystal structure of catechol *O*-methyltransferase. *Nature* **368**, 354–358.
6. Fu, Z., *et al.*, & Takusagawa, F. (1996). Crystal structure of glycine *N*-methyltransferase from rat liver. *Biochemistry* **35**, 11985–11993.
7. Park, I.K. & Paik, W.K. (1990). The occurrence and analysis of methylated amino acids. In *Protein Methylation*. (Paik, W.K. & Kim, S., eds), pp. 1–22, CRC Press, Inc, Boca Raton.
8. Clarke, S. (1985). Protein carboxyl methyltransferase: two distinct classes of enzymes. *Annu. Rev. Biochem.* **54**, 479–506.
9. Ota, I.M. & Clarke, S. (1990). The function and enzymology of protein D-aspartyl/L-aspartyl methyltransferases in eukaryotic and prokaryotic cells. In *Protein Methylation*. (Paik, W.K. & Kim, S., eds), pp. 179–194, CRC Press, Inc, Boca Raton.
10. Clarke, S. (1992). Protein isoprenylation and methylation at carboxy-terminal cysteine residues. *Ann. Rev. Biochem.* **61**, 355–386.
11. Philips, M.R., *et al.*, & Stock, J.B. (1993). Carboxyl methylation of Ras-related proteins during signal transduction in neutrophils. *Science* **259**, 977–980.
12. Springer, W.R. & Koshland, D.E., Jr. (1977). Identification of a protein methyltransferase as the *cheR* gene product in the bacterial sensing system. *Proc. Natl. Acad. Sci. USA* **74**, 533–537.
13. Springer, M.S., Goy, M.F. & Adler, J. (1979). Protein methylation in behavioral control mechanisms and in signal transduction. *Nature.* **280**, 279–284.
14. Koshland, D.E., Jr. (1988). Chemotaxis as a model second-messenger system. *Biochemistry* **27**, 5829–5834.
15. Stock, J.B. & Surette, M.G. (1996). Chemotaxis. In *Escherichia coli and Salmonella: cellular and molecular biology.* (Neidhardt, F.C., *et al.*, & Umbarger, H.E., eds), pp. 1103–1129, ASM Press, Washington DC, USA.
16. Stewart, R.C. & Dahlquist, F.W. (1988). N-terminal half of CheB is involved in methylesterase response to negative chemotactic stimuli in *Escherichia coli*. *J. Bacteriol.* **170**, 5728–5738.
17. Lupas, A. & Stock, J. (1989). Phosphorylation of an N-terminal regulatory domain activates the CheB methylesterase in bacterial chemotaxis. *J. Biol. Chem.* **264**, 17337–17342.

18. Stock, J.B. & Koshland, D.E., Jr. (1981). Changing reactivity of receptor carboxyl groups during bacterial sensing. *J. Biol. Chem.* **256**, 10826–10833.
19. Springer, M.S., Zanolari, B. & Pierzchala, P.A. (1982). Ordered methylation of the the methyl-accepting chemotaxis proteins of *Escherichia coli*. *J. Biol. Chem.* **257**, 6861–6866.
20. Kehry, M.R., Doak, T.G. & Dahlquist, F.W. (1984). Stimulus-induced changes in methylesterase activity during chemotaxis in *Escherichia coli*. *J. Biol. Chem.* **259**, 11828–11835.
21. Shapiro, M.J. & Koshland, D.E., Jr. (1994). Mutagenic studies of the interaction between the aspartate receptor and methyltransferase from *Escherichia coli*. *J. Biol. Chem.* **269**, 11054–11059.
22. Shapiro, M.J., Chakrabarti, I. & Koshland, D.E., Jr. (1995). Contributions made by individual methylation sites of the *Escherichia coli* aspartate receptor to chemotactic behavior. *Proc. Natl. Acad. Sci. USA* **92**, 1053–1056.
23. Wu, J., Li, J., Li, G., Long, D.G. & Weis, R.M. (1996). The receptor binding site for the methyltransferase of bacterial chemotaxis is distinct from the sites of methylation. *Biochemistry* **35**, 4984–4993.
24. Malone, T., Blumenthal, R.M. & Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253**, 618–632.
25. Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. (1995). Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**, 16–20.
26. Holm, L. & Sander, C. (1993). Dali Version 2.0. *J. Mol. Biol.* **233**, 123–138.
27. West, A.H., Martinez-Hackert, E. & Stock, A.M. (1995). Crystal structure of the catalytic domain of the chemotaxis receptor methylesterase, CheB. *J. Mol. Biol.* **250**, 276–290.
28. Simms, S.A. & Subbaramaiah, K. (1991). The kinetic mechanism of *S*-adenosyl-L-methionine: glutamylmethyltransferase from *Salmonella typhimurium*. *J. Biol. Chem.* **266**, 12741–12746.
29. Branden, C.I., *et al.*, & Akeson, A. (1973). Structure of liver alcohol dehydrogenase at 2.9-angstrom resolution. *Proc. Natl. Acad. Sci. USA* **70**, 2439–2442.
30. Subbaramaiah, K., Charles, H. & Simms, S.A. (1991). Probing the role of cysteine residues in the CheR methyltransferase. *J. Biol. Chem.* **266**, 19023–19027.
31. Subbaramaiah, K. & Simms, S.A. (1992). Photolabeling of CheR methyltransferase with *S*-adenosyl-L-methionine (AdoMet). *J. Biol. Chem.* **267**, 8636–8642.
32. Ingrosso, D., Fowler, A.V., Bleibaum, J. & Clarke, S. (1989). Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes: common sequence motifs for protein, DNA, RNA, and small molecule *S*-adenosylmethionine-dependent methyltransferases. *J. Biol. Chem.* **264**, 20131–20139.
33. O'Gara, M., McCloy, K., Malone, T. & Cheng, X. (1995). Structure-based sequence alignment of three AdoMet-dependent DNA methyltransferases. *Gene* **157**, 135–138.
34. Schluckebier, G., Kozak, M., Bleimling, N., Weinhold, E. & Saenger, W. (1997). Differential binding of *S*-Adenosylmethionine *S*-Adenosylhomocysteine and Sinefungin to the adenine-specific DNA methyltransferase M, *Taq*I. *J. Mol. Biol.* **265**, 56–67.
35. Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. (1994). Hhal methyltransferase flips its target base out of the DNA helix. *Cell.* **76**, 357–369.
36. Kehry, M.R., Engstrom, P., Dahlquist, F.W. & Hazelbauer, G.L. (1983). Multiple covalent modifications of Trg, a sensory transducer of *Escherichia coli*. *J. Biol. Chem.* **258**, 5050–5055.
37. Terwilliger, T.C., Bogonez, E., Wang, E.A. & Koshland, D.E., Jr. (1983). Sites of methyl esterification on the aspartate receptor involved in bacterial chemotaxis. *J. Biol. Chem.* **258**, 9608–9611.
38. Nowlin, D.M., Bollinger, J. & Hazelbauer, G.L. (1987). Sites of covalent modification in Trg, a sensory transducer of *Escherichia coli*. *J. Biol. Chem.* **262**, 6039–6045.
39. Rice, M.S. & Dahlquist, F.W. (1991). Sites of deamidation and methylation in Tsr, a bacterial chemotaxis sensory transducer. *J. Biol. Chem.* **266**, 9746–9753.
40. Terwilliger, T.C. & Koshland, D.E., Jr. (1984). Sites of methyl-esterification and deamination on the aspartate receptor involved in chemotaxis. *J. Biol. Chem.* **259**, 7719–7725.
41. Lupas, A., VanDyke, M. & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164.
42. Shapiro, M.J., Panomitros, D. & Koshland, D.E., Jr. (1995). Interactions between the methylation sites of *Escherichia coli* aspartate receptor mediated by the methyltransferase. *J. Biol. Chem.* **270**, 751–755.
43. Stock, J.B., Clarke, S. & Koshland, D.E., Jr. (1984). The protein carboxylmethyltransferase involved in *Escherichia coli* and *Salmonella typhimurium* chemotaxis. *Methods Enzymol.* **106**, 310–321.
44. Long, D.G. & Weis, R.M. (1992). *Escherichia coli* aspartate receptor. Oligomerization of the cytoplasmic fragment. *Biophys. J.* **62**, 69–71.
45. Floss, H.G., Mascaro, L., Tsai, M.-D. & Woodard, R.W. (1979). Stereochemistry of enzymatic transmethylation. In *Transmethylation*. (Usdin, E., Borchardt, R.T. & Creveling, C.R., eds), pp. 135–141, Elsevier North Holland Inc., New York, USA.
46. Nicholls, A. & Honig, B. (1991). A rapid finite difference algorithm, utilizing successive over relaxation to solve the Poisson–Boltzmann equation. *J. Comput. Chem.* **12**, 435–445.
47. West, A.H., Djordjevic, S., Martinez-Hackert, E. & Stock, A.M. (1995). Purification, crystallization, and preliminary X-ray diffraction analyses of the bacterial chemotaxis receptor modifying enzymes. *Proteins* **21**, 345–350.
48. Otwinowski, Z. (1993). Oscillation data reduction program. *In Proceedings of the CCP4 Study Weekend: Data Collection and Processing.* (Sawyer, L., Isaacs, N. & Bayley, S., eds), pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.
49. SERC (UK) Collaborative Computational Project, No.4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760–763.
50. Furey, W. & Swaminathan, S. (1997). PHASES-95: a program package for the processing and analysis of diffraction data from macromolecules. *Methods Enzymol.* **377**, in press.
51. Otwinowski, Z. (1991). Isomorphous replacement and anomalous scattering. In *Proceedings of the CCP4 Study Weekend.* (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80–86, SERC Daresbury Laboratory, Warrington, U.K.
52. Cowtan, K. (1994). 'dm': an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter Protein Crystallography* **31**, 34–38.
53. Roussel, A. & Inisan, A.G. (1992). TURBO-manual, Bio-Graphics. Marseille, France.
54. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460.
55. Lamzin, V.S. & Wilson, K.S. (1993). Automated refinement of protein models. *Acta Cryst. D* **49**, 129–147.
56. Carson, M. (1991). Ribbons 2.0. *J. Appl. Cryst.* **24**, 958–961.
57. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
58. Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387–395.

2702–2715 Nucleic Acids Research, 1997, Vol. 25, No. 14

Structure of *Pvu*II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment

Weimin Gong, Margaret O'Gara, Robert M. Blumenthal[1] and Xiaodong Cheng*

W.M.Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA and [1]Department of Microbiology and Immunology, Medical College of Ohio, Toledo, OH 43699-0008, USA

Received May 7, 1997; Revised and Accepted May 27, 1997

ABSTRACT

We have determined the structure of *Pvu*II methyltransferase (M.*Pvu*II) complexed with *S*-adenosyl-L-methionine (AdoMet) by multiwavelength anomalous diffraction, using a crystal of the selenomethionine-substituted protein. M.*Pvu*II catalyzes transfer of the methyl group from AdoMet to the exocyclic amino (N4) nitrogen of the central cytosine in its recognition sequence 5′-CAGCTG-3′. The protein is dominated by an open α/β-sheet structure with a prominent V-shaped cleft: AdoMet and catalytic amino acids are located at the bottom of this cleft. The size and the basic nature of the cleft are consistent with duplex DNA binding. The target (methylatable) cytosine, if flipped out of the double helical DNA as seen for DNA methyltransferases that generate 5-methylcytosine, would fit into the concave active site next to the AdoMet. This M.*Pvu*II α/β-sheet structure is very similar to those of M.*Hha*I (a cytosine C5 methyltransferase) and M.*Taq*I (an adenine N6 methyltransferase), consistent with a model predicting that DNA methyltransferases share a common structural fold while having the major functional regions permuted into three distinct linear orders. The main feature of the common fold is a seven-stranded β-sheet (6↓ 7↑ 5↓ 4↓ 1↓ 2↓ 3↓) formed by five parallel β-strands and an antiparallel β-hairpin. The β-sheet is flanked by six parallel α-helices, three on each side. The AdoMet binding site is located at the C-terminal ends of strands β1 and β2 and the active site is at the C-terminal ends of strands β4 and β5 and the N-terminal end of strand β7. The AdoMet–protein interactions are almost identical among M.*Pvu*II, M.*Hha*I and M.*Taq*I, as well as in an RNA methyltransferase and at least one small molecule methyltransferase. The structural similarity among the active sites of M.*Pvu*II, M.*Taq*I and M.*Hha*I reveals that catalytic amino acids essential for cytosine N4 and adenine N6 methylation coincide spatially with those for cytosine C5 methylation, suggesting a mechanism for amino methylation.

INTRODUCTION

DNA methyltransferases (Mtases) transfer a methyl group from *S*-adenosyl-L-methionine (AdoMet) to a given position of a particular DNA base within a specific DNA sequence. The resulting methylation can protect the DNA from a cognate restriction endonuclease or can have epigenetic effects on gene expression. The DNA Mtases belong to two families: one methylates C5, a ring carbon of cytosine, yielding 5-methylcytosine (5mC), while the second family methylates the exocyclic amino group (NH_2) of cytosine or adenine yielding *N*4-methylcytosine (N4mC) or *N*6-methyladenine (N6mA) respectively. Two of the 5mC Mtases have been structurally characterized as covalent reaction intermediate complexes with their DNA substrates (1,2); one of these, M.*Hha*I has been characterized in complexes with structural analogs of DNA in three different methylation states, unmethylated, hemimethylated and fully methylated (3,4).

The primary sequences of the 5mC Mtases share a set of conserved motifs (I–X) in a constant linear order (5–9). The majority of these motifs are responsible for three basic functions of the 5mC Mtases: AdoMet binding, sequence-specific DNA binding and catalysis of methyl transfer. In contrast, the amino-Mtases (which generate N6mA or N4mC) belong to three groups characterized by distinct linear orders for the conserved motifs (10). The three groups are named α (including Mtases such as Dam), β (including Mtases such as M.*Pvu*II) and γ (including Mtases such as M.*Taq*I). To date only one DNA amino-Mtase has been structurally characterized, the group γ N6mA Mtase M.*Taq*I (11).

While the M.*Taq*I structure has been determined only in the absence of DNA, it is sufficient to allow general structural comparison with the 5mC Mtases. Both M.*Hha*I and M.*Taq*I are bilobal structures: one lobe contains a catalytic domain with both the active site for methyl transfer and the AdoMet binding site and the other lobe contains a target (DNA) recognition domain (TRD). The catalytic domains of the two proteins exhibit very

*To whom correspondence should be addressed at present address: Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA. Tel: +1 404 727 8491; Fax: +1 404 727 3746; Email: xcheng@emory.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

similar three-dimensional folding (12). This folding pattern is also present in M.*Hae*III, another 5mC Mtase, in catechol *O*-Mtase, a single domain small molecule AdoMet-dependent Mtase, in VP39, an mRNA cap-specific RNA 2′-*O*-Mtase and in glycine *N*-Mtase (2,13–15). The folding similarity includes the positions of conserved amino acid side chains involved in either AdoMet binding or catalysis; only the binding of AdoMet reported for glycine *N*-Mtase differs from the consensus pattern (15). Guided by this common catalytic domain structure, sequence alignment of amino-Mtases suggests that for all amino-Mtases to fit the consensus M.*Hha*I/M.*Taq*I catalytic domain structure, despite having different motif orders, different sets of topological connections would be required for the three DNA amino-Mtase groups (10).

Determining the structure of *Pvu*II methyltransferase (M.*Pvu*II), a group β N4mC Mtase, would thus address two important questions about DNA Mtases. First, do the N4mC Mtases in fact match the consensus catalytic domain structure seen between M.*Taq*I and M.*Hha*I (12)? Second, are the major structural elements of amino-Mtases connected in three different orders, as suggested by their primary sequences (10)? M.*Taq*I itself did not provide a strong test for this model because the group γ and 5mC Mtases have essentially the same motif order; they differ only in the position of motif X (10). No Mtase from group α or β has been structurally characterized before this report.

M.*Pvu*II, part of the restriction–modification system from the Gram-negative bacterium *Proteus vulgaris* (16), modifies the internal cytosine of the recognition sequence 5′-CAGCTG-3′ (17) to generate N4mC (18). *Pvu*II endonuclease, which cleaves duplex DNA at the center of the same recognition sequence to generate blunt-ended products, was structurally characterized earlier (19,20). With this report, the *Pvu*II restriction–modification system becomes the first system for which the structures of the cognate endonuclease and methyltransferase have both been determined.

MATERIALS AND METHODS

Overexpression and crystallization

Overexpression and purification of and selenomethionine (SeMet) incorporation into M.*Pvu*II have been described previously (21). To crystallize the M.*Pvu*II–AdoMet binary complex, 0.2 mM AdoMet was added to the pre-purified protein (~5 μM) and the mixture was further purified by cation exchange chromatography (21). M.*Pvu*II and selenomethionyl M.*Pvu*II, complexed with AdoMet, both crystallized in the monoclinic space group $P2_1$ with unit cell dimensions of $a = 48.8$ Å, $b = 112.4$ Å, $c = 59.3$ Å and $\beta = 109.2°$ (21). There are two molecules per crystallographic asymmetric unit cell, termed molecules A and B. X-Ray diffraction data were collected using a MarResearch imaging plate detector on beamline X12-C at the National Synchrotron Light Source, Brookhaven National Laboratory, and processed using the HKL software package (22). Multiwavelength anomalous diffraction (MAD; 23) data to 3.3 Å resolution (Table 1) were collected on a single frozen SeMet crystal at three different wavelengths, corresponding to the inflection point λ1 (minimum $\Delta f'$) and the peak λ2 (maximum $\Delta f''$) of the Se-containing crystal absorption spectrum and a third wavelength (λ3) remote from the peak position (21). A higher resolution data set (up to 2.8 Å) used for final model refinement was collected from a native crystal (λ4 = 1.072 Å, 180° rotation, 1.5° increment, 90 s exposure).

SeMet MAD phasing

There are a total of 18 possible Se sites per asymmetric unit (nine per molecule). To locate the Se positions, we calculated the anomalous and isomorphous difference Patterson maps at the Harker section ($v = 1/2$) among data sets collected at wavelengths λ1, λ2 and λ3. A number of peaks were observed, which corresponded to possible Se sites and the cross vectors between them (21). Five Se sites were first manually determined from the Patterson maps. These five sites were used to calculate initial estimates of phases, to compute the difference and Bijvoet difference Fourier synthesis and to search for additional Se sites. Finally, a total of 12 Se sites were determined and confirmed by the two-fold non-crystallographic symmetry (NCS) operator, revealed by a self-rotation function (21).

Table 1. Statistics of experimental SeMet MAD data with rejection criteria $I/\sigma(I) \geq 2$

	λ1	λ2	λ3
Wavelength (Å)	0.98233	0.98211	0.92
Energy (eV)	12 621	12 624	13 476
Resolution range (Å)	∞–3.30		
Completeness (%)	94.5	92.2	94.5
$R_{linear} = \Sigma\|I - \langle I\rangle\|/\Sigma I$	0.048	0.051	0.044
$\langle I/\sigma\rangle$	17.0	16.1	18.6
Observed reflections	29 236	27 108	27 410
Unique reflections	8 904	8 790	8 913
Anomalous pairs	7 576	8 212	7 475
Highest resolution shell (Å)	3.36–3.30		
Completeness (%)	90.0	87.5	90.5
$R_{linear} = \Sigma\|I - \langle I\rangle\|/\Sigma I$	0.097	0.101	0.087
$\langle I/\sigma\rangle$	9.6	8.9	11.1
Unique reflections	425	391	431

These 12 Se positions were used for MAD phasing by treating the data from each wavelength as a multiple isomorphous replacement experiment with the inclusion of anomalous scattering (MIRAS): native with native anomalous scattering (λ3), derivative isomorphous (λ1) and derivative isomorphous with anomalous scattering (λ2; Table 2). The MAD–MIRAS phases were improved using 40% solvent content by four, four and eight cycles of solvent leveling (24) following each of three envelope determinations (25). The solvent-leveled map was used to refine the NCS operator and to construct the averaging mask. The phases were further improved using 16 rounds of Furey's averaging protocols (25). The electron density was averaged within the mask, the density for each molecule was replaced with the average and the 'averaged' density map was inverted to obtain new phases. The resulting phases were combined with the original solvent-leveled

MAD-MIRAS phases. The process was cycled until convergence was obtained (16 cycles).

Refinement

The starting Cα backbone for molecule A was traced using the skeleton in program O (26) with reference to three maps at 3.3 Å resolution: MAD-MIRAS, solvent flattened and density averaged. The Se positions also provide markers for selenomethionine in the polypeptide chain tracing. The atomic coordinates for molecule B were generated by the two-fold NCS operator. After the initial model building, the atomic model was subjected to refinement against 2.8 Å resolution data from a native crystal (Table 3). Initially, a strict NCS was invoked, assuming that two NCS-related molecules are strictly identical. The two models were refined by simulated annealing and least squares minimizations using the X-PLOR program suite (27). Seven rounds of refinement and model rebuilding brought the crystallographic R factor to 0.22. The model was further refined by a restraint NCS, with two NCS-related atoms restrained in their average positions. An additional five rounds of refinement, refitting and placing ordered water molecules brought the R factor to 0.19 and R_{free} to 0.28 (Table 3).

Table 2. Treatment of SeMet MAD data as MIRAS at 3.3 Å resolution

Native wavelength		λ3	
Derivative wavelength	λ1	λ2	λ3
isomorphous/anomalous	iso	iso/ano	ano
Phasing power[a]	2.350	1.830/1.410	1.080
R-Kraut[b]	0.028	0.028/0.037	0.029
R-Cullis[c]	0.516	0.601/–	–
Figure of merit	0.389	0.365/0.308	0.265
Overall figure of merit	0.619		

[a]Phasing power = r.m.s.($\langle F_H \rangle / E$), where F_H is the calculated 'heavy atom' structure factor amplitude and E is the residual lack of closure.
[b]R-Kraut = $\Sigma|F_{PH} - F_{PH}^{cal}|/\Sigma|F_{PH}|$ for centric data and $\Sigma(|F_{PH+} - F_{PH+}^{cal}| + |F_{PH-} - F_{PH-}^{cal}|)/\Sigma|F_{PH+} + F_{PH-}|$ for acentric data.
[c]R-Cullis = $\Sigma||F_{PH} \pm F_P| - F_H|/\Sigma|F_{PH} \pm F_P|$ for centric data, where F_{PH} and F_P are the observed structure factor amplitudes for the 'derivative' and 'native' data sets and F_H is the calculated 'heavy atom' structure factor amplitude.

RESULTS

Structure determination

M.*Pvu*II is produced in two forms, resulting from translation initiators 13 codons apart (17). The shorter form of M.*Pvu*II, starting from the internal translation initiator at Met14, was overexpressed in *Escherichia coli* and purified both in native and selenomethionine-substituted forms (21). The M.*Pvu*II polypeptide chain is 323 amino acids long (numbered 14–336). Diffraction data (Table 1) were collected at three X-ray wavelengths from a crystal of the selenomethionyl M.*Pvu*II–AdoMet complex, so that MAD could be used to extract the phases (23). Following the suggestion of Ramakrishnan (28,29), multiwavelength data were treated as if they were from a conventional MIRAS experiment. A total of 12 (out of 18 possible) Se sites per asymmetric unit were determined from Patterson maps and were used for MAD phasing with a figure of merit of 0.62 at 3.3 Å resolution. The MAD-MIRAS map, coupled with two-fold non-crystallographic symmetry averaging, was accurate enough to permit an initial interpretation (21) and the model was finally refined to 2.8 Å resolution with a crystallographic R factor of 0.19 and an R_{free} value of 0.28.

Table 3. Structural refinement of M.*Pvu*II at 2.8 Å resolution

Wavelength (Å)	1.072	
Resolution range (Å)	∞–2.8	2.85–2.8 (highest resolution shell)
Completeness (%)	99.7	97.5
$R_{linear} = \Sigma|I - \langle I \rangle|/\Sigma I$	0.052	0.226
$\langle I/\sigma \rangle$	11.5	5.5
Observed reflections	54 787	
Unique reflections	14 886	730
R_{factor}[a]$= \Sigma|F_o - F_c|/\Sigma|F_o|$	0.193	
R_{free}[a]	0.283	
Non-hydrogen protein atoms	4455	
r.m.s. deviation from ideality		
Bond lengths (Å)	0.02	
Bond Angles (°)	2.9	
Dihedrals (°)	24.9	
Improper (°)	2.2	

[a]R_{factor} and R_{free} are calculated for ~92 and 8% of the data respectively.

Overview of the M.*Pvu*II structure

The polypeptide chain folds into a structure with a V-shaped cleft, big enough to accommodate duplex DNA (Fig. 1). The V-shaped cleft is formed by three loops on one side and a three-helix bundle on the other side. The methyl donor AdoMet binds at the bottom of the cleft, which consists of a twisted 10 stranded β-sheet around which six α-helices are arranged on both sides.

Figure 2 shows the topology diagrams of M.*Pvu*II, M.*Hha*I and M.*Taq*I. For clarity and convenience, we retain the nomenclature of Schluckebier *et al.* (12) for the secondary structure assignment and of Posfai *et al.* (5) for the conserved motifs. Loops or turns are designated by their flanking secondary structures; two of them are termed the glycine-containing G loop (loop 1-A) and the proline-containing P loop (loop 4-D) (10). The catalytic domains of the three structures are all of the α/β type with a central β-sheet sandwiched between two layers of α-helices: helices αC, αD and αE located on one side and helices αZ, αA and αB on the opposite side of the sheet (Fig. 2a). The β-sheets in the three structures all contain five central adjacent parallel β-strands with strand order 5, 4, 1, 2, 3 and one antiparallel hairpin (β6 and β7) next to strand β5. The order of parallel strands is reversed once between β4 and β1. The majority of the active amino acids from conserved motifs (circled in Fig. 2) are located at the carboxyl ends or in loop regions outside the carboxyl ends of these parallel β-strands. In all three structures the AdoMet binding site is located at the carboxyl ends of strands β1 and β2 and the amino end of helix αC;

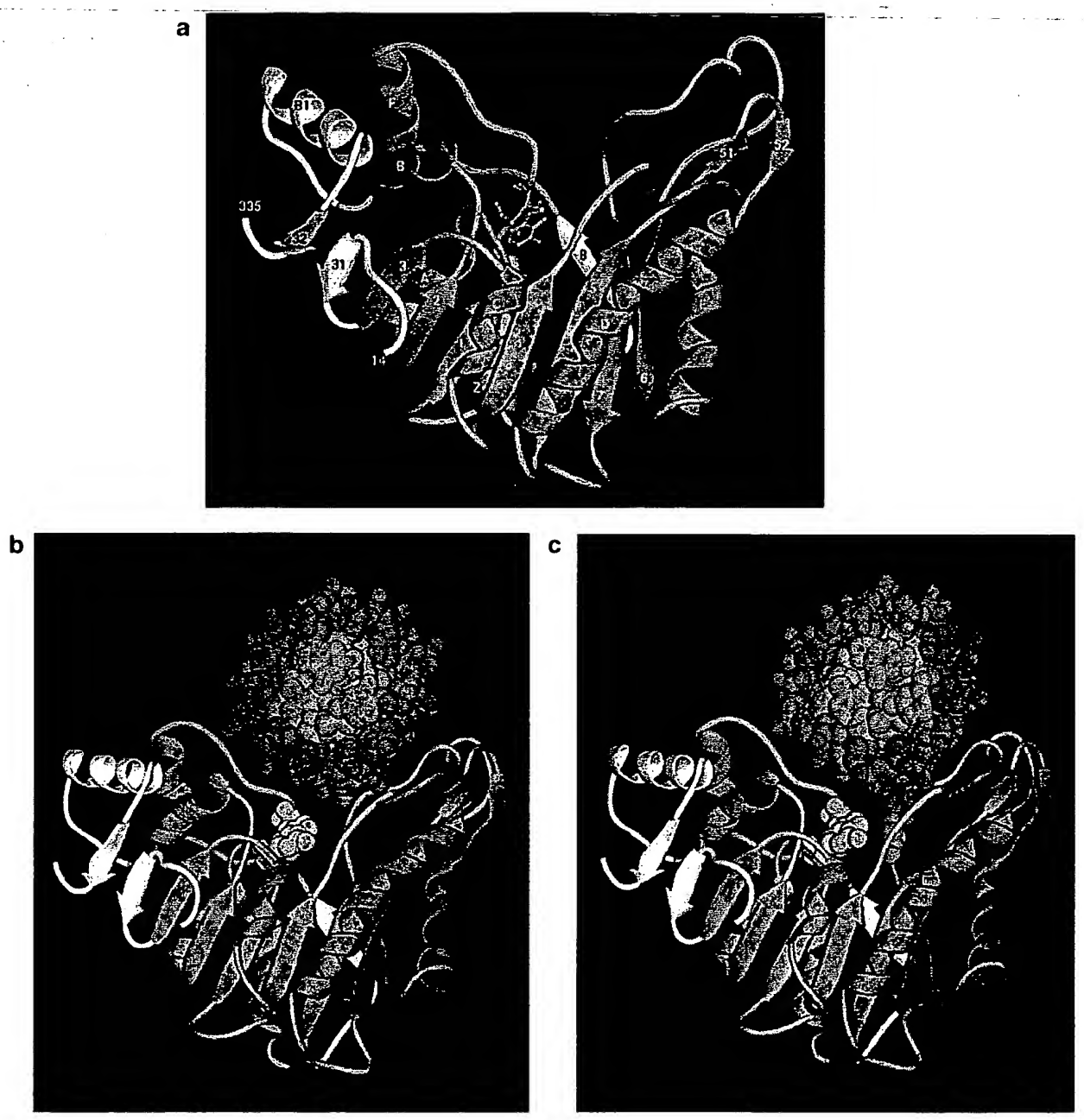


Figure 1. Ribbon (70) diagram of M.*Pvu*II. (**a**) The protein folds into a structure with a V-shaped cleft. AdoMet (in ball-and-stick representation) is bound at the bottom of the cleft. The regions of M.*Pvu*II that are structurally most similar to M.*Hha*I and M.*Taq*I are shown in brown and the less similar regions are shown in white and green. The green region is part of the putative DNA target recognition domain (TRD). The catalytic P loop is in pale blue and red. The pale blue part contains conserved amino acids Ser53–Pro–Pro–Phe56 and the red part is flexible (high thermal factors), consistent with potential conformational change upon DNA binding. (**b**) M.*Pvu*II docked to cognate DNA, taken from the R.*Pvu*II–DNA structure (19). The DNA phosphate backbone and sugar rings are in purple, the DNA bases are in green and AdoMet is in yellow. (**c**) M.*Pvu*II docked to DNA with a flipped cytosine (see Fig. 5).

and the active site at the carboxyl ends of strands β4 and β5 and the amino end of the strand β7 (see below). The N- and C-termini of the folded polypeptide are within the AdoMet binding region in all three structures: located in the region between helix αZ and strand β1 (M.*Hha*I in Fig. 2b), prior to helix αZ (M.*Taq*I in Fig. 2c) and between helix αB and strand β3 (M.*Pvu*II in Fig. 2d).

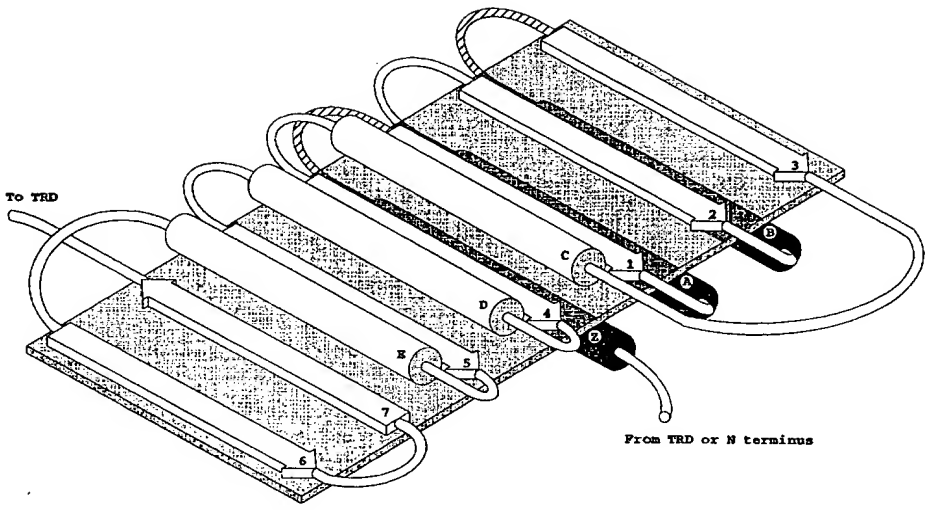


Figure 2. Topology diagrams. (a) The consensus methylase fold, derived from DNA Mtases (M.*Hha*I, M.*Hae*III, M.*Taq*I and M.*Pvu*II) and one small molecule AdoMet-dependent Mtase (catechol *O*-Mtase). The main feature of the fold is a region of five parallel β-strands (5, 4, 1, 2, 3) followed by an antiparallel β-hairpin (strands 6 and 7), surrounded by six helices, three (αC, αD, αE and αZ, αA, αB) on each side of the β-sheet. The dashed loops can be broken to become the N- and C-termini. Opposite page: (b) M.*Hha*I, (c) M.*Taq*I and (d) M.*Pvu*II diagrams indicate their similarity in the catalytic domains. α-Helices are shown as rectangles (lettered) and β-strands as broad arrows (numbered). Common elements of secondary structure among the three enzymes are shown in similar positions. Conserved or functionally important amino acids from motifs I–X are circled. The β-strands (ten in M.*Pvu*II, seven in M.*Hha*I and nine in M.*Taq*I) form a β-sheet. In M.*Pvu*II, a region (dashed line) between strands β7 and β8 is not modeled in the current structure. (e) Predicted topological folding for group β amino-Mtases from an earlier study (10).

M.*Pvu*II fits the consensus fold for AdoMet-dependent Mtases

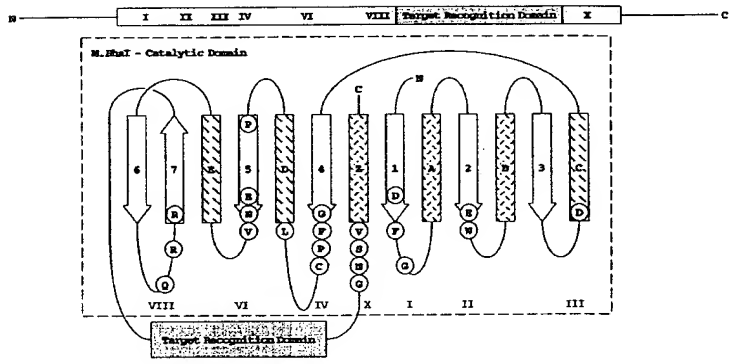
The TRD, which is associated with sequence-specific DNA recognition, lies in the smaller domain of the three bilobal Mtases discussed above. In the current structure of M.*Pvu*II, this domain comprises only one helix (αF) and its associated loops (Fig. 2d). It is interesting how the TRD is connected to the catalytic domain in the three Mtases. In 5mC Mtases such as M.*Hha*I, helix αZ (motif X, part of the catalytic domain) is folded from the C-terminus, following the TRD. Thus, there are two connections between the catalytic domain and TRD (Fig. 2b). In group γ N6mA Mtase M.*Taq*I, helix αZ originates from the N-terminus and the TRD is linked to the catalytic domain through β9 only (Fig. 2c). Thus, in both M.*Taq*I and M.*Hha*I the functional regions are in the order (amino→carboxyl) AdoMet binding region, active site region and TRD, the major difference between them being that helix αZ is moved from the N-terminus in M.*Taq*I to the C-terminus in M.*Hha*I.

As predicted (10), the most pronounced difference in topology between M.*Pvu*II and both M.*Hha*I and M.*Taq*I is the connection between the AdoMet binding and active site regions: the two regions are connected via the putative TRD (helix αF) in the order (amino→carboxyl) active site region, TRD and AdoMet binding region (Fig. 2d). The active site and AdoMet binding regions of M.*Pvu*II fit the consensus structure of M.*Hha*I/M.*Taq*I/M.*Hae*III/ catechol *O*-Mtase, regardless of the motif order in the primary sequence. We call this common catalytic domain structure the AdoMet-dependent methylase fold (Fig. 2a). This fold has also been observed in the RNA Mtase VP39, though helix αE is replaced by a β-strand (14).

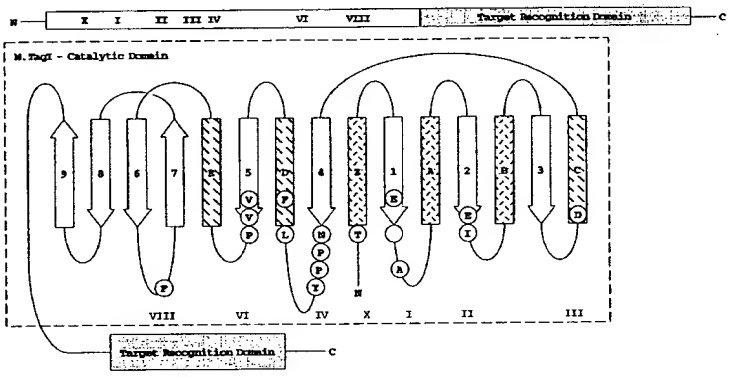
We had predicted the folding of group β amino-Mtases, including M.*Pvu*II (Fig. 2e), based on structure-guided sequence analysis (10). Overall, the prediction is quite accurate, though there are some significant differences between the prediction and the current model. Unexpectedly, part of the AdoMet binding region (β3–αC or motif III) is located upstream of the active site region, near the N-terminus of the polypeptide. This arrangement preserves the crossover between strands β3 and β4, but splits the coding for the AdoMet binding region into two distant parts of the gene. The β3–αC secondary structure (motif III) was predicted to originate from the C-terminus as a contiguous part of the AdoMet binding region. This prediction, which would result in no crossover connection between strands β3 and β4, was made in part because of the very short distance between the N-terminus of another group β Mtase (M.*Bam*HII) and its strand β4 (Fig. 3). However, this crossover has been observed in all currently available DNA Mtase structures (5mC Mtases M.*Hha*I and M.*Hae*III, group β Mtase M.*Pvu*II and group γ Mtase M.*Taq*I), as well as in catechol *O*-Mtase, glycine *N*-Mtase and the RNA Mtase VP39, and is predicted to occur in group α amino-Mtase structures, with the crossover connection in a separate domain comprising the TRD (10).

Such a crossover is necessary to generate a so-called topological switch point (30), at which the strand order is reversed and loops connected to the carboxyl ends of the two adjacent strands (β1 and β4 in Fig. 2a) go in opposite directions. The positions of concave active sites can be predicted from such switch points in different types of α/β twisted open sheet structures, including arabinose

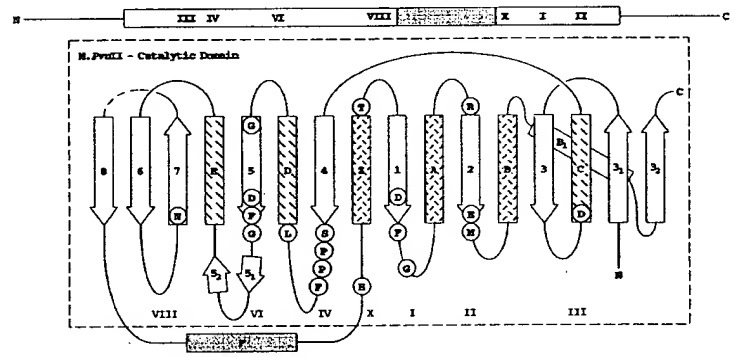
b


N
I XI XII IV VI VIII Target Recognition Domain X
C
M.HaeI - Catalytic Domain
Target Recognition Domain
VIII VI IV X I II III

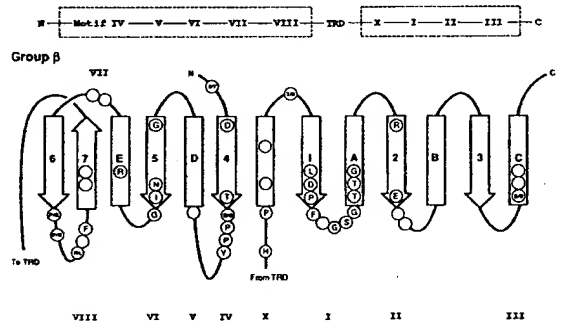
c


N
X I XI XII IV VI VIII Target Recognition Domain
C
M.TaqI - Catalytic Domain
Target Recognition Domain
VIII VI IV X I II III

d


N
XII IV VI VIII X I XI
C
M.PvuII - Catalytic Domain
VIII VI IV X I II III

e


N
Motif IV V VI VII VIII TRD X I II III
C
Group β
To TRD
From TRD
VIII VI V IV X I II III

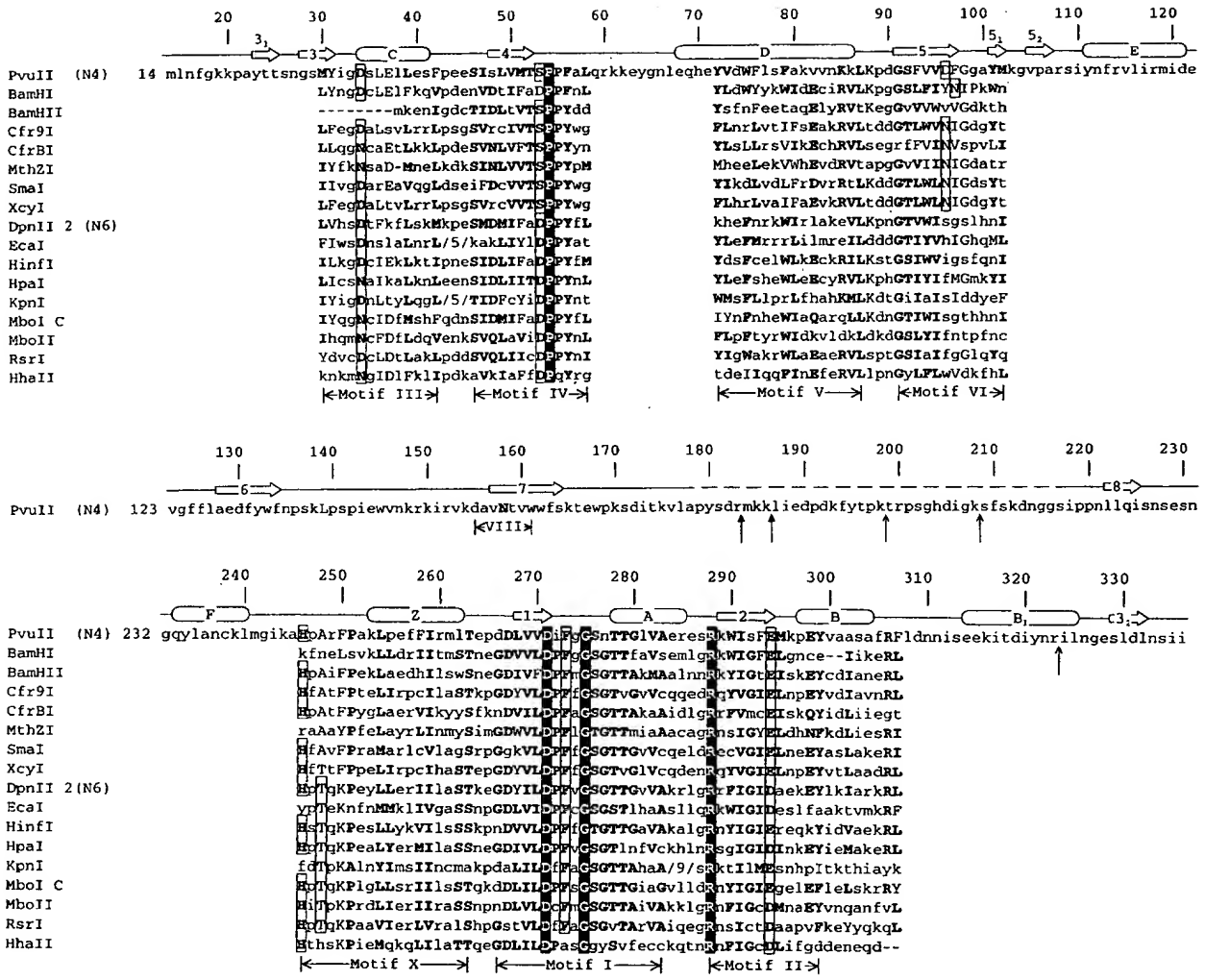


Figure 3. Sequence alignment of group β amino-Mtases including eight N4mC and nine N6mA Mtases (there is some uncertainty on assignments, particularly for M.*Hin*fI). Conserved amino acids are grouped as (E, D, Q, N), (V, L, I, M), (F, Y, W), (G, P, A), (K, R) and (S, T), using standard one letter abbreviations. Invariant amino acids are shown as white letters against a black background; conserved positions are indicated by bold letters within a box. Lesser degrees of conservation are shown, in decreasing order, by bold and upper case letters, while non-conserved positions are shown as lower case letters. A dash (–) indicates a deletion relative to other sequences and a slash (/) followed by a number indicates an insertion and its size. Motifs I–X are labeled using the nomenclature of Posfai *et al.* (5). The secondary structures of M.*Pvu*II are indicated by cylinders (α-helices) and arrows (β-strands) drawn directly above the amino acids forming them. The dashed line indicates a flexible region (amino acids 179–216) which is not modeled in the current structure. This region includes four out of five preferred trypsin cleavage sites indicated by arrows (34).

binding protein (31), carboxypeptidase (32) and tyrosyl-tRNA synthetase (33).

Disordered regions

As noted above, there are two molecules per crystallographic asymmetric unit cell, termed molecules A and B. The current model of molecule A contains residues a16–a178, a217–a335 and one AdoMet, while molecule B contains residues b16–b56, b69–b178, b215–b335 and one AdoMet. The r.m.s. deviation between 269 common Cα atoms of the final refined two molecules is 0.6 Å. In both molecules ~40 amino acids (Pro179–Gly216), located immediately after strand β7 and before strand β8, were not modeled in the current structure because of poor electron density. This poor density suggests that these amino acids are very flexible. Consistent with this flexibility, four out of five preferred trypsin cleavage sites are within this 40 amino acid region: the primary cleavages occur on the carboxyl sides of Arg183 and Lys186 and are followed by slower cleavages carboxyl of Lys198, Lys208 and Arg323 (34). In fact, SDS–PAGE analysis of dissolved crystals indicates that some M.*Pvu*II crystals contained limited amounts of protein that had been cleaved in this region. It is noteworthy in this regard that some 5mC Mtases are naturally made as two separate polypeptides that associate in the cell to form active enzyme (35,36).

In molecule B part of the catalytic P loop (amino acids 57–68) was also not modeled due to poor electron density. However the corresponding P loop in molecule A was modeled, though Leu58–Asn66 (red in Fig. 1) possessed the highest crystallographic thermal factors in the current refined structure. This flexibility may be due to the absence of the DNA in the crystal and suggests a potential conformational change upon DNA binding. Similarly, the catalytic P loop in M.*Hha*I, which contains the key catalytic amino acids Pro80–Cys81, undergoes a massive conformational change upon binding DNA, moving ~25 Å toward the corresponding DNA binding cleft of the protein (1).

AdoMet binding

The binding site for AdoMet is adjacent to the carboxyl ends of strands β1, β2, the amino end of helix αC and the loop prior to helix αZ, regions that contain conserved motifs I, II, III and X respectively (Fig. 4). The interactions between AdoMet and M.*Pvu*II are almost identical to those between AdoMet and M.*Hha*I (1), M.*Taq*I (11), catechol *O*-Mtase (13) and VP39 (14). Amino acid side chains interacting with AdoMet are found in spatially equivalent positions, except that Phe273 of M.*Pvu*II and Phe18 of M.*Hha*I are in the G loop, while the corresponding Phe146 of M.*Taq*I is in helix αD (Fig. 2).

In motif I of group β Mtases (Asp–X–Phe–X–Gly), the amino acids Asp and Gly are invariant. In M.*Pvu*II these correspond to Asp271 and Gly275 (Fig. 3). The side chain carboxylate of Asp271 (β1) makes two hydrogen bonds to the main chain amide group of Phe273 (G loop) and the side chain hydroxyl of Thr279 (αA) and these bonds stereochemically constrain the β1–loop–αA structure. A negatively charged amino acid corresponding to Asp271 has been found in the same position of motif I in all DNA Mtases sequenced so far, including Asp16 of M.*Hha*I and Glu45 of M.*Taq*I (9,10). The main chain amide group of Gly275 (G loop) hydrogen bonds to the side chain carboxylate of Glu294 (β2), which is another conserved negatively charged amino acid (motif II) that interacts with the ribose hydroxyls of AdoMet. Comparable backbone–side chain interactions occur in M.*Hha*I (Gly20–Glu40) and M.*Taq*I (Ala49–Glu71).

In M.*Pvu*II the AdoMet binding α/β cluster (αZ→β1→αA→β2→αB) is further stabilized by the interactions of Arg288 (an invariant arginine among group β Mtases located prior to strand β2; Fig. 3) with the side chain of Thr263 (loop Z-1) and backbone carboxyls of both Thr263 (loop Z-1) and Glu286 (loop A-2). Only three structurally characterized AdoMet binding proteins interact with AdoMet in substantially different ways from the nucleic acid Mtases and catechol *O*-Mtase. One of these proteins is the *E.coli* MetJ repressor (37), for which AdoMet is a co-repressor and not a substrate; another is the reactivation domain of *E.coli* methionine synthase (38), which uses AdoMet in a flavodoxin-coupled reductive methylation of cobalamin. The third is glycine *N*-Mtase which does have a region structurally very similar to the consensus AdoMet-binding regions, though that is not where AdoMet was bound in the reported structure (15).

AdoMet binding and target base binding sites are structurally similar to one another

The M.*Pvu*II protein has approximate two-fold pseudo symmetry around the center of the cleft, due in part to the structural similarity of the AdoMet binding site to the active site. These sites are each dominated by comparable α/β clusters, αZ→β1→αA→β2→αB and αC→β4→αD→β5→αE; the former includes motifs I, II and X and forms the bulk of the AdoMet binding region and the latter includes motifs IV–VI and forms the bulk of the active site region. The two α/β clusters can be superimposed by rotating strands β1 and β2 onto strands β4 and β5 (Fig. 4b). This yields an r.m.s. deviation of 0.7 Å for the Cα atoms of these β-strands. Similar superimposability has also been observed for the α/β clusters of the 5mC Mtases M.*Hha*I and M.*Hae*III and the N6mA Mtase M.*Taq*I (10). This observation has led to the suggestion that the original Mtases arose after gene duplication converted an AdoMet binding protein into a protein that bound two molecules of AdoMet (see also 39–42) and that the two halves then diverged (10). Regardless of the evolutionary model, the M.*Pvu*II structure suggests that this internal structural repeat is a feature common to most AdoMet-dependent Mtases. Only the reactivation domain of *E.coli* methionine synthase does not fit this pattern (38).

DISCUSSION

Predicted DNA binding and base flipping

It is very likely that the V-shaped cleft of the protein is where DNA binds. In the absence of large scale protein conformational changes, the cleft is large enough to accommodate double-stranded DNA without steric hindrance (Fig. 1b). Positively charged groups, capable of interacting with the DNA phosphate backbone, are prominent on the surface of the cleft from the P loop (Arg60–Lys–Lys62), loop 5-E (Lys103 and Arg108) and loops 6–7 (Lys138, Lys148–Arg–Lys150, Arg152 and Lys154). We have docked a 13mer B-DNA duplex containing the *Pvu*II recognition sequence, taken from the R.*Pvu*II–DNA structure (19), against the basic face of the cleft (Fig. 1b). The fit of the DNA in the cleft is extremely convincing, with the protein occupying a distance of ~37 Å along the axis of the double helix, which suggests that M.*Pvu*II intimately contacts a 10 nt stretch including the 6 nt recognition sequence.

The M.*Hha*I–DNA structure provided the first example of base flipping (1). Several other types of enzymes are now also known or believed to use this approach (43,44), including the DNA repair enzymes T4 endonuclease V and human uracil-DNA glycosylase (45,46). The M.*Pvu*II structure is consistent with a base flipping mechanism. Base flipping is a process by which an enzyme can rotate a DNA nucleotide out of the double helix, breaking only the base pairing hydrogen bonds and trapping it in a protein binding pocket. In our docking model the DNA is positioned such that the target cytosine is in the helix and the NH_2 group to be methylated is far from the active CH_3 group of AdoMet (Fig. 1b). Thus it is likely that M.*Pvu*II (an amino-Mtase) flips the cytosine out of the DNA helix to access the target amino group (Fig. 1c), in a manner similar to that employed by 5mC Mtases, M.*Hha*I and M.*Hae*III (1–4). The structure of M.*Taq*I and spectroscopic data for M.*Eco*RI suggest that these two amino-Mtases flip the target adenine out of DNA (47,48).

Although it is possible to predict where DNA binds, we cannot identify any known DNA binding motifs in the current structure that might be responsible for DNA sequence specificity. Furthermore, there is no obvious similarity between M.*Pvu*II and the structures of R.*Pvu*II or MyoD in complex with DNA (19,49); both are homodimeric proteins recognizing the same DNA sequence, CAGCTG. R.*Pvu*II uses a β-ribbon motif to interact with nucleotides in the DNA major groove, while the myogenic

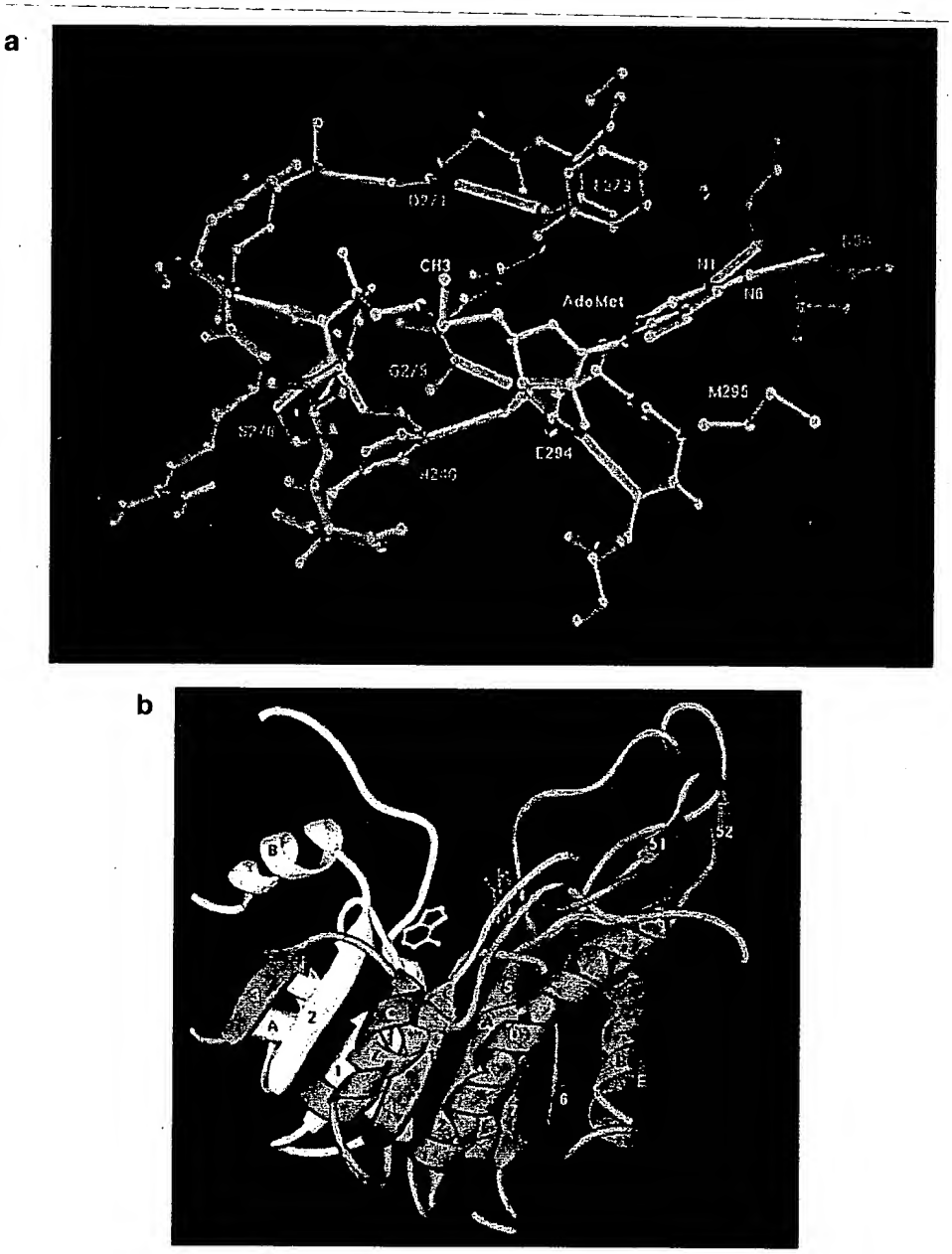


Figure 4. AdoMet binding site. (a) AdoMet is involved in contacts with four regions, formed by four motifs, the G loop (motif I in brown), strand β (motif II in pale blue), helix αC (motif III in purple) and loop F-Z (motif X in green). (b) Superimposition of the two α/β clusters. The first cluster, αZ→β1→αA→β2→αB, in white, is rotated (now in green) with respect to the second cluster, αC→β4→αD→β5→αE, in brown, to achieve the most overlap possible. Also shown are the positions, relative to the respective α/β clusters, of the AdoMet adenosyl moiety (white and green) and the target cytosine ring, in brown (inferred from M.*Hha*I–DNA structure, see Fig. 5).

transcription factor MyoD is a basic helix–loop–helix protein. The lack of obvious similarity may reflect the disparate roles of these three CAGCTG-recognizing proteins. DNA Mtases carry out base flipping (within specific nucleotide sequences) so they can access the atom to be methylated on the target nucleotide. Such a mechanism is not required for other sequence-specific proteins, such as transcription factors (for which specific binding is the main role) and restriction endonucleases (which only act on the readily accessible DNA phosphate backbone).

As mentioned before, only two 5mC Mtases, M.*Hha*I and M.*Hae*III, have been structurally characterized in complex with their DNA substrates. The protein–base contacts in the recognized sequence are expected to differ between M.*Hha*I and M.*Hae*III due to their different specificity and, indeed, the folding of the

corresponding TRDs is different (2). However, both TRDs contain a shared feature: two recognition loops (1,2,44). In the M.*Pvu*II structure, two loops (prior to and after helix αF) on the other side of the V-shaped cleft could easily fit into the concave face of the major or minor groove of B-form DNA. These two loops, which may correspond to the two 5mC recognition loops, are held in place through scaffolding made up of three helices, αF, αB and αB_1. A similar pair of recognition loops has also been proposed for M.*Taq*I (47). The reason for such conservation may be that sequence recognition is a part of the base flipping mechanism and loops, instead of the more rigid structures of α-helix or β-strand, are used for discriminating DNA sequences flexibly and effectively.

Predicted catalytic mechanism for DNA amino methylation

What we call the catalytic P loop of the amino-Mtases was found in early sequence comparisons and called an 'Asp–Pro–Pro–Tyr motif' based on its sequence (50,51). A later comparison suggested it might correspond to Pro–Cys (motif IV) in 5mC Mtases, even though the reaction mechanisms of the two families of Mtases appear to be quite distinct (52). The structural comparison of M.*Hha*I and M.*Taq*I has confirmed that the Pro–Cys and Asn–Pro–Pro–Tyr motifs of these two enzymes are spatially equivalent (12) and thus, by analogy, are referred to as motif IV (10). Motif IV has the consensus sequence Ser–Pro–Pro–Tyr for N4mC Mtases, Asp–Pro–Pro–Tyr for groups α and β N6mA Mtases and Asn–Pro–Pro–Tyr for group γ N6mA Mtases (10,53,54). However, as we discuss below, Ser→Asp→Asn must not present an essential functional difference. We note that these consensus sequences are not absolute and there is still a problem in distinguishing N4mC from N6mA Mtases just on the basis of amino acid sequence (see Fig. 3).

The flipped cytosine, taken from the M.*Hha*I–DNA structure, can be docked surprisingly well into the M.*Pvu*II active site, located at the bottom of the V-shaped cleft. By superimposition of the common α/β-sheet structures, the active site amino acids in M.*Hha*I from the catalytic P loop and strands β5 and β7 overlap the corresponding amino acids in M.*Pvu*II: Gly78–Phe–Pro–Cys81 onto Ser53–Pro–Pro–Phe56 (P loop), Glu119 onto Asp96 (β5) and Arg165 onto Asn158 (β7) (Fig. 5a). In M.*Hha*I these amino acids interact with the target cytosine: Arg165 interacts with O2, Glu119 with N3 and N4, the main chain carbonyl of Phe79 with N4 and Cys81 covalently bonds to C6. Though M.*Pvu*II also interacts with cytosine, we do not observe the identical amino acids in the same structural elements in M.*Pvu*II. However, as noted above, different amino acids are spatially equivalent in the two enzymes.

One can easily model the interactions between the polar edge of the flipped cytosine and M.*Pvu*II (shown in brown in Fig. 5a). The target atom, cytosine N4, would have two possible hydrogen bond partners: the hydroxyl group of Ser53 and the main chain carbonyl of Pro54 (the first two amino acids of the highly conserved motif IV). Also from this conserved motif, the phenyl ring of Phe56 could make van der Waals contacts with the cytosine ring; Phe56 occupies a position similar to Cys81 of M.*Hha*I. Asn158 (β7), which does not appear to be conserved among the N4mC Mtases, might hydrogen bond to cytosine O2.

Asp96 (Asn in most of the other N4mC Mtases) may hydrogen bond with and activate the Ser53 hydroxyl group (Asp96:$O_{\delta 2}$...Ser53:O_γ = 2.7 Å), thereby facilitating proton transfer from the cytosine amino group through the Ser and eventually to the Asp (Fig. 6a). If this occurs, the protonated Asp96 might then hydrogen bond to the N3 of the cytosine. Ser53 and Asp96 thus appear to belong to a charge relay system analogous to that seen in the serine proteinases (55).

Most importantly, the distance of the AdoMet methyl group to the cytosine N4 is ~4 Å in our docking model, sufficiently close to permit methyl group transfer. For comparison, in the structures of M.*Hha*I–DNA complexes the substrate cytosine C5–AdoMet methyl distance is ~2.9 Å (56); the product 5mC methyl–AdoHcy sulfur is also ~2.9 Å (4). Thus, our model suggests that methylation of the exocyclic amino group results from a direct attack of the activated cytosine N4 on the AdoMet methyl group, in analogy with the previously proposed mechanism for DNA adenine methylation (12,57,58).

In the group β N4mC Mtases, Ser53 of M.*Pvu*II is conserved except in M.*Bam*HI, which has Asp at this position (Fig. 3); a conserved Asp is present in the same place in the group β N6mA Mtases, as well as in the group α N6mA Mtases (10). Modeling suggests that Asp in this position of the P loop could interact with cytosine N4 and N3 (M.*Bam*HI) or adenine N6 and N1 (Fig. 6b). The Asp carboxyl group could hydrogen bond with cytosine N4 (NH_2) or adenine N6 (NH_2), thereby increasing the nucleophilicity of the nitrogen and serving as a trap for the amino-leaving proton, when the methyl group transfers to the nitrogen from AdoMet. In that case the protonated carboxyl group could hydrogen bond with cytosine N3 or adenine N1. If this is correct, the conserved Asp in M.*Bam*HI and the N6mA Mtases may be functionally comparable to Asp96 in M.*Pvu*II. Ser53 in M.*Pvu*II may compensate for the fact that Asp96 is too far from the cytosine N4 for direct interaction (Figs 5a and 6a), but this does not explain why most N4mC Mtases do not simply have Asp in place of Ser, as is seen in M.*Bam*HI.

When the structures of M.*Pvu*II and M.*Taq*I, a group γ N6mA Mtase, are superimposed at their common α/β-sheet structures, Asn105 of Asn–Pro–Pro–Tyr (P loop) in M.*Taq*I is present in place of Ser53 of Ser–Pro–Pro–Phe in M.*Pvu*II; and two hydrophobic amino acids, Phe196 (loops 6–7) and Val163 (β5), of M.*Taq*I replace the positions of two polar/charged groups (Asn158 and Asp96) of M.*Pvu*II. These hydrophobic amino acids, particularly Phe196, are likely to make van der Waals contacts with the target nucleotide (12). The carboxamide of M.*Taq*I Asn105 could interact with both adenine N6 and N1 (Fig. 6c), similar to the role Asn229 of thymidylate synthase plays in hydrogen bonding to dUMP (see figure 3 of 59). However, Asn229 of thymidylate synthase plays a contributory but non-essential role in catalysis (60). In contrast to Asp or Ser, it is unlikely that Asn can accept a proton. Therefore, the only role obvious at the present time played by the Asn of the group γ N6mA Mtases is in positioning the substrate adenine, while the methylation would result from a direct attack of the AdoMet methyl group on the adenine N6 with a general base (possibly a highly ordered water molecule) assisting the proton transfer that occurs at N6.

A second AdoMet molecule

Some Mtases, including M.*Pvu*II, appear to bind two molecules of AdoMet (34,61,62), one of which affects the selectivity of the protein towards substrate and non-specific DNA sequences (for M.*Eco*Dam; 61,62). Extra electron density (from $2F_o - F_c$, $F_o - F_c$

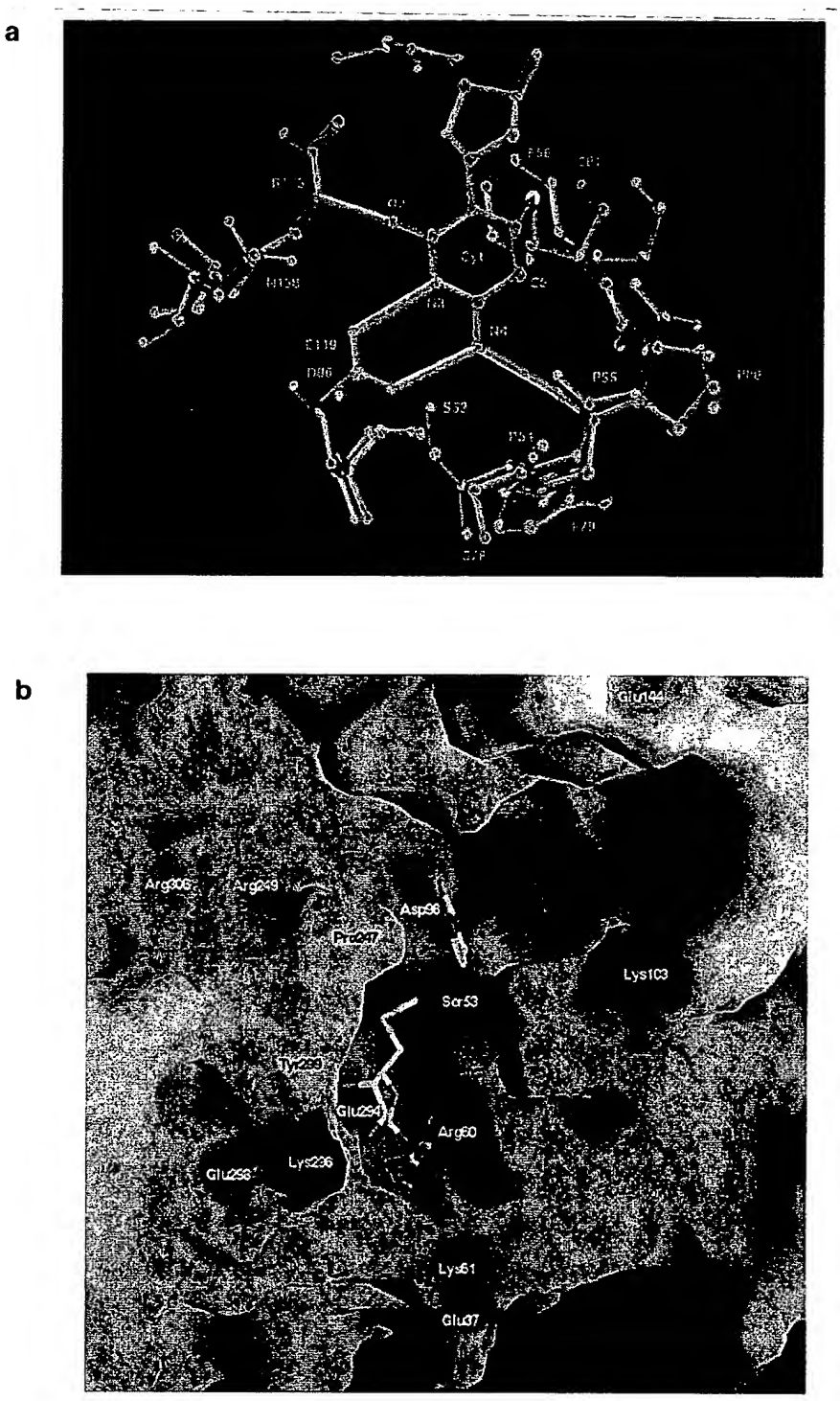


Figure 5. Active site. (a) Superimposition of the active sites in M.*Pvu*II (brown) and M.*Hha*I (green). Amino acids shown are from the P loop (motif IV) and strands β5 (motif VI) and β7 (motif VIII) (see Fig. 2). In the complex between M.*Hha*I and a transition state analog substrate, Cys81 is linked by a covalent bond (yellow) to C6 of the target cytosine (1). The cytosine is recognized by a number of hydrogen bonds (in white). (b) Close-up GRASP (71) representation displayed at the level of the solvent accessible surface. Color coded purple for positive (~20 K_BT), red for negative (—20 K_BT) and white for neutral, where K_B is the Boltzmann's constant and T is the temperature. The AdoMet and the modeled target cytosine ring are in stick representation, with yellow for carbon, purple for nitrogen, red for oxygen, green for sulfur. The second AdoMet binding site is formed by the first AdoMet molecule at the bottom, Tyr299 and Pro247 on the left side and the main chains of Pro55 and Phe56 (between Ser53 and Arg60) on the right side.

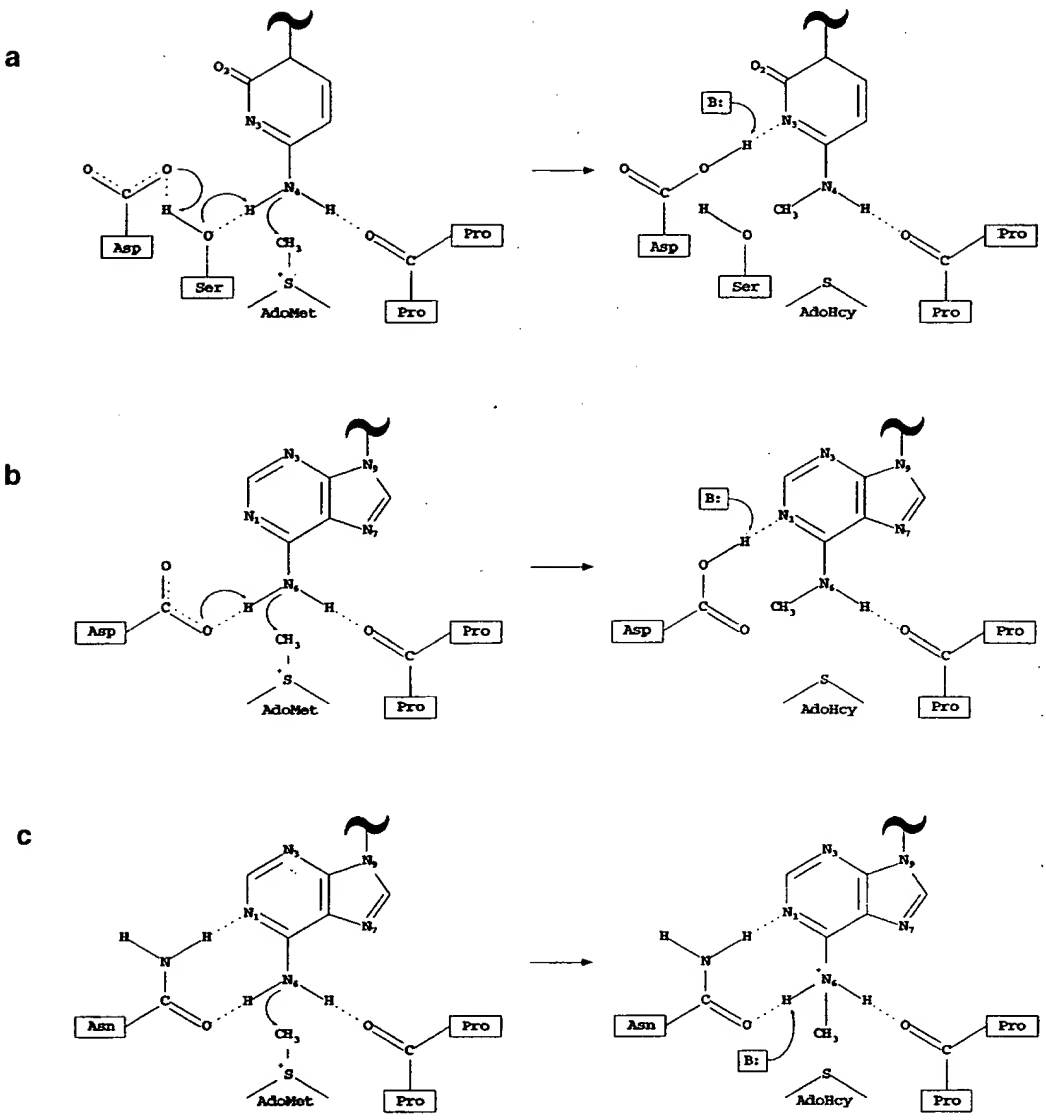


Figure 6. Proposed reaction mechanisms for amino-Mtases with the P loop containing (a) Ser (as in M.*Pvu*II), (b) Asp and (c) Asn. A general base (B:), which could be a water molecule, might be needed to eliminate the proton.

and initial MAD-MIRAS maps) was found near the first AdoMet in molecule A. This may be a second AdoMet, as the density can be fitted well to an AdoMet adenosyl moiety with the methionine moiety extending into the solvent. This second AdoMet binding site is formed by the first AdoMet molecule at the bottom, Tyr299 (αB) and His246–Pro247 (loop F-Z) on one side and the main chains of Pro55 and Phe56 (P loop) on the other (Fig. 5b). The adenine sits above the ribose ring of the first AdoMet. Most interestingly, the second AdoMet ribose oxygens interact with the side chain of Glu37 of crystallographic symmetry-related molecule B. This interaction, analogous to the first AdoMet–Glu294 (β2, motif II), may stabilize the second AdoMet in molecule A, due to the different crystal packing environment.

However, despite the structural similarity of the AdoMet binding and active sites (Fig. 4b), this second AdoMet molecule does not occupy the active site (Fig. 5b). Instead, the second AdoMet occupies a space equivalent to the solvent channel in the M.*Hha*I–DNA structure, where a network of well-ordered water molecules, including that proposed as the general base for eliminating the C5 proton, mediates contacts between the target cytosine, AdoHcy and M.*Hha*I (see figure 3 of 4).

Evolutionary relationships among the DNA Mtases

As noted above, the structure of M.*Pvu*II confirms two predicted features of DNA Mtase structure. First, all DNA Mtases

structurally characterized to date have AdoMet adenine binding pockets that are superimposable onto their methylatable base binding pockets (10; see Fig. 4b). Second, all DNA Mtases structurally characterized to date share a common α/β architecture for their catalytic domains, making different topological connections to accommodate the permuted linear orders of functional regions in their genes (see Fig. 2).

These two features have implications for models of the evolutionary relationships among DNA Mtases. The internal symmetry provided by the two binding pockets, each formed by a comparable set of α helices and β strands, is suggestive of evolution by gene duplication (10). Subsequent gene fusion could have converted the resulting small molecule Mtase to a DNA Mtase by adding a TRD; some DNA Mtases are still produced in two separate pieces that associate to form active enzyme and one piece is essentially the TRD while the other is the catalytic domain (35,36).

The second feature, common structure despite permuted gene orders, raises a question. Do the four groups of DNA Mtases (α, β, γ and 5mC) represent divergence from a common ancestor or convergence from separate Mtase lineages? Matthews *et al.* (63) have proposed a set of six criteria for distinguishing divergence from convergence: the DNA sequences of the genes should be similar, the amino acid sequences of the proteins should be similar, the three-dimensional structures should be similar, the enzyme–substrate interactions should be similar, the catalytic mechanisms should be similar and '...those segments of the polypeptide chain that are critical for catalysis are in the same sequence in the respective proteins (i.e. insertions and deletions are allowed, but not transpositions)'. There is as yet no structure for a Mtase of the α group, but Mtases from the other three groups (where the information is known) satisfy all except the last criterion (Fig. 2): the DNA Mtase groups have the major functional regions in three permuted gene orders (10). We can only note that several proteins have been found to remain structurally and functionally intact following circular permutation of their genes (64–68) and that genetic mechanisms for gene permutation have been proposed (69). Whether convergence or divergence describes the relationship between the DNA Mtases, it is clear that the N4mC Mtases such as M.*Pvu*II do not represent a separate subfamily of enzymes.

NOTE ADDED IN PROOF

Since acceptance of this paper, the structure of an. AdoMet-dependent protein methyltransferase has been published (72). The *Salmonella typhimurium* CheR protein matches the consensus Mtase structure very well, including the binding of AdoHcy in the expected AdoMet pocket.

ACKNOWLEDGEMENTS

We acknowledge Robert M.Sweet for help with X-ray data collection at the Brookhaven National Laboratory, in the Biology Department single-crystal diffraction facility, at beamline X12-C in the National Synchrotron Light Source. That facility is supported by the US Department of Energy, Office of Health and Environmental Research and by the National Science Foundation. We thank Thomas Malone for preparing Figures 2, 3 and 6, Rowena G.Matthews for critical discussions, Richard J.Roberts and David T.F.Dryden for comments on the manuscript and Kim Gernert for preparing the cover figure. This report is partially funded by a National Institutes of Health fellowship (GM17052 to M.O'G.), the National Science Foundation (MCB-9631137 to R.M.B.) and the National Institutes of Health (GM49245 and GM/OD52117 to X.C.).

REFERENCES

1 Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) *Cell*, **76**, 357–369.
2 Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) *Cell*, **82**, 143–153.
3 O'Gara,M., Klimasauskas,S., Roberts,R.J. and Cheng,X. (1996) *J. Mol. Biol.*, **261**, 634–645.
4 O'Gara,M., Roberts,R.J. and Cheng,X. (1996) *J. Mol. Biol.*, **263**, 597–606.
5 Posfai,J., Bhagwat,A.S., Posfai,G. and Roberts,R.J. (1989) *Nucleic Acids Res.*, **17**, 2421–2435.
6 Lauster,R., Trautner,T.A. and Noyer-Weidner,M. (1989) *J. Mol. Biol.*, **206**, 305–312.
7 Som,S., Bhagwat,A.S. and Friedman,S. (1987) *Nucleic Acids Res.*, **15**, 313–332.
8 Cheng,X., Kumar,S., Posfai,J., Pflugrath,J.W. and Roberts,R.J. (1993) *Cell*, **74**, 299–307.
9 Kumar,S., Cheng,X., Klimasauskas,S., Mi,S., Posfai,J., Roberts,R.J. and Wilson,G.G. (1994) *Nucleic Acids Res.*, **22**, 1–10.
10 Malone,T., Blumenthal,R.M. and Cheng,X. (1995) *J. Mol. Biol.*, **253**, 618–632.
11 Labahn,J., Granzin,J., Schluckebier,G., Robinson,D.P., Jack,W.E., Schildkraut,I. and Saenger,W. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 10957–10961.
12 Schluckebier,G., O'Gara,M., Saenger,W. and Cheng,X. (1995) *J. Mol. Biol.*, **247**, 16–20.
13 Vidgren,J., Svensson,L.A. and Liljas,A. (1994) *Nature*, **368**, 354–358.
14 Hodel,A.E., Gershon,P.D., Shi,X. and Quiocho,F.A. (1996) *Cell*, **85**, 246–257.
15 Fu,Z., Hu,Y., Konishi,K., Takata,Y., Ogawa,H., Gomi,T., Fujioka,M. and Takusagawa,F. (1996) *Biochemistry*, **35**, 11985–11993.
16 Gingeras,T.R., Greenough,L., Schildkraut,I. and Roberts,R.J. (1981) *Nucleic Acids Res.*, **9**, 4525–4536.
17 Blumenthal,R.M., Gregory,S.A. and Cooperider,J.S. (1985) *J. Bacteriol.*, **164**, 501–509.
18 Butkus,V., Klimasauskas,S., Petrauskiene,L., Maneliene,Z., Lebionka,A. and Janulaitis,A.A. (1987) *Biochim. Biophys. Acta*, **909**, 201–207.
19 Cheng,X., Balendiran,K., Schildkraut,I. and Anderson,J.E. (1994) *EMBO J.*, **13**, 3927–3935.
20 Athanasiadis,A., Vlassi,M., Kotsifaki,D., Tucker,P.A., Wilson,K.S. and Kokkinidis,M. (1994) *Nature Struct. Biol.*, **1**, 469–475.
21 O'Gara,M., Adams,G.M., Gong,W., Kobayashi,R., Blumenthal,R.M. and Cheng,X. (1997) *Eur. J. Biochem.*, in press.
22 Otwinowski,Z. (1993). In Sawyer,L., Issacs,N. and Bailey,S (eds), *Data Collection and Processing*. SERC Daresbury Laboratory, Warrington, UK, pp. 56–62.
23 Hendrickson,W.A. (1991) *Science*, **254**, 51–58.
24 Wang,B.C. (1985) *Methods Enzymol.*, **115**, 90–112.
25 Furey,W. and Swaminathan,S. (1996) *Methods Enzymol.*, **277**, in press
26 Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgard,M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
27 Brünger,A.T. (1992) *X-PLOR. A System for X-ray Crystallography and NMR*, version 3.1. Yale University, New Haven, CT.
28 Ramakrishnan,V., Finch,J.T., Graziano,V., Lee,P.L. and Sweet,R.M. (1993) *Nature*, **362**, 219–223.
29 Ramakrishnan,V. and Biou,V. (1996) *Methods Enzymol.*, **276A**, 538–557.
30 Brändén,C.-I. (1980) *Q. Rev. Biophys.*, **13**, 317–338.
31 Gilliland,G.L. and Quiocho,F.A. (1981) *J. Mol. Biol.*, **146**, 341–362.
32 Rees,D.C., Lewis,M. and Lipscomb,W.N. (1983) *J. Mol. Biol.*, **168**, 367–387.
33 Brick,P., Bhat,T.N. and Blow,D.M. (1988) *J. Mol. Biol.*, **208**, 83–98.
34 Adams,G.M. and Blumenthal,R.M. (1997) *Biochemistry*, in press.
35 Karreman,C. and de Waard,A. (1990) *J. Bacteriol.*, **172**, 266–272.
36 Lee,K.-F., Kam,K.-M. and Shaw,P.-C. (1995) *Nucleic Acids Res.*, **23**, 103–108.
37 Somers,W.S. and Phillips,S.E.V. (1992) *Nature*, **359**, 387–393.

38 Dixon,M.M., Huang,S., Matthews,R.G. and Ludwig,M. (1996) *Structure*, **4**, 1263–1275.
39 Lauster,R. (1988) *Gene*, **74**, 243.
40 Lauster,R. (1989) *J. Mol. Biol.*, **206**, 313–321.
41 Tao,T., Walter,J., Brennan,K.J., Cotterman,M.M. and Blumenthal,R.M. (1989) *Nucleic Acids Res.*, **17**, 4161–4175.
42 Guyot,J.-B. and Caudron,B. (1994) *C.R. Acad Sci. Paris III Sci. Vie.*, **317**, 20–24.
43 Roberts,R.J. (1995) *Cell*, **82**, 9–12.
44 Cheng,X. and Blumenthal,R.M. (1996) *Structure*, **4**, 639–645.
45 Vassylyev,D.G., Kashiwagi,T., Mikami,Y., Ariyoshi,M., Iwai,S., Ohtsuka,E. and Morikawa,K. (1995) *Cell*, **83**, 773–782.
46 Slupphaug,G., Mol,C.D., Kavli,B., Arvai,A.S., Krokan,H.E. and Tainer,J.A. (1996) *Nature*, **384**, 87–92.
47 Schluckebier,G., Labahn,J., Granzin,J., Schildkraut,I. and Saenger,W. (1995) *Gene*, **157**, 131–134.
48 Allan,B.W. and Reich,N.O. (1996) *Biochemistry*, **35**, 14757–14762.
49 Ma,P.C., Rould,M.A., Weintraub,H. and Pabo,C.O. (1994) *Cell*, **77**, 451–459.
50 Hattman,S., Wilkinson,J., Swinton,D., Schlagman,S., Macdonald,P.M. and Mosig,G. (1985) *J. Bacteriol.*, **164**, 932–937.
51 Chandrasegaran,S. and Smith,H.O. (1988) In Sarma,R.H. and Sarma,M.H. (eds), *Structure and Expression: from Proteins to Ribosomes*. Adenine Press, New York, NY, Vol. 1, pp. 149–156.
52 Klimasauskas,S., Timinskas,A., Menkevicius,S., Butkiene,D., Butkus,V. and Janulaitis,A. (1989) *Nucleic Acids Res.*, **17**, 9823–9832.
53 Wilson,G.G. (1992) *Methods Enzymol.*, **216**, 259–279.
54 Wilson,G.G. and Murray,N.E. (1991) *Annu. Rev. Genet.*, **25**, 585–627.
55 Warshel,A., Naray-Szabo,G., Sussman,F. and Hwang,J.K. (1989) *Biochemistry*, **28**, 3629–3637.
56 Kumar,S., Horton,J.H., Jones,G.D., Walker,R.T., Roberts,R.J. and Cheng,X. (1997) *Nucleic Acids Res.*, **25**, 2773–2783.
57 Pogolotti,A.L., Ono,A., Subramaniam,R. and Santi,D.V. (1988) *J. Biol. Chem.*, **263**, 7461–7464.
58 Ho,D.K., Wu,J.C., Santi,D.V. and Floss,H.G. (1991) *Arch. Biochem. Biophys.*, **284**, 264–269.
59 Gerlt,J.A. (1994) *Curr. Opin. Struct. Biol.*, **4**, 593–600.
60 Liu,L. and Santi,D.V. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 8604–8608.
61 Bergerat,A. and Guschlbauer,W. (1990) *Nucleic Acids Res.*, **18**, 4369–4375.
62 Bergerat,A., Guschlbauer,W. and Fazakerley,G.V. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 6394–6397.
63 Matthews,B.W., Remington,S.J., Grhtter,M.G. and Anderson,W.F. (1981) *J. Mol. Biol.*, **147**, 545–558.
64 Luger,K., Hommel,U., Herold,M., Hofsteenge,J. and Kirschner,K. (1989) *Science*, **243**, 206–210.
65 Buchwalder,A., Szadkowski,H. and Kirschner,K. (1992) *Biochemistry*, **31**, 1621–1630.
66 Hahn,M., Piotukh,K., Borriss,R. and Heinemann,U. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 10417–10421.
67 Ritco-Vonsovici,M., Minard,P., Desmadril,M. and Yon,J.M. (1995) *Biochemistry*, **34**, 16543–16551.
68 Graf,R. and Schachman,H.K. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 11591–11596.
69 Gilbert,W. (1985) *Science*, **228**, 823–824.
70 Carson,M. (1991) *J. Appl. Crystallogr.*, **24**, 958–961.
71 Nicholls,A., Sharp,K.A. and Honig,B. (1991) *Proteins Struct. Funct. Genet.*, **11**, 281–296.
72 Djordjevic,S. and Stock,A.M. (1997) *Structure*, **5**, 545–558.

Express Mail No.
EV335394187US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants	:	Marc Pignot et al.
Application No.	:	09/744,641
Filed	:	January 26, 2001
For	:	NEW COFACTORS FOR METHYLTRANSFERASES

Examiner	:	Josephine Young
Art Unit	:	1645
Docket No.	:	630182.401USPC
Date	:	July 23, 2004

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION UNDER 37 C.F.R. § 1.132

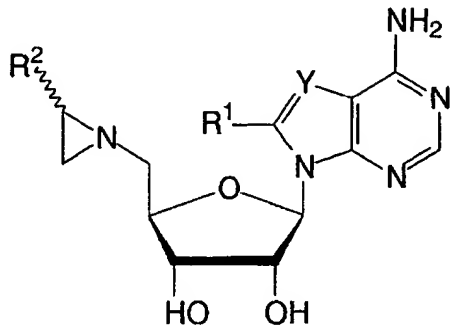
Sir:

1. I, Elmar Weinhold, am a co-inventor with Marc Pignot, on the above-identified patent application.

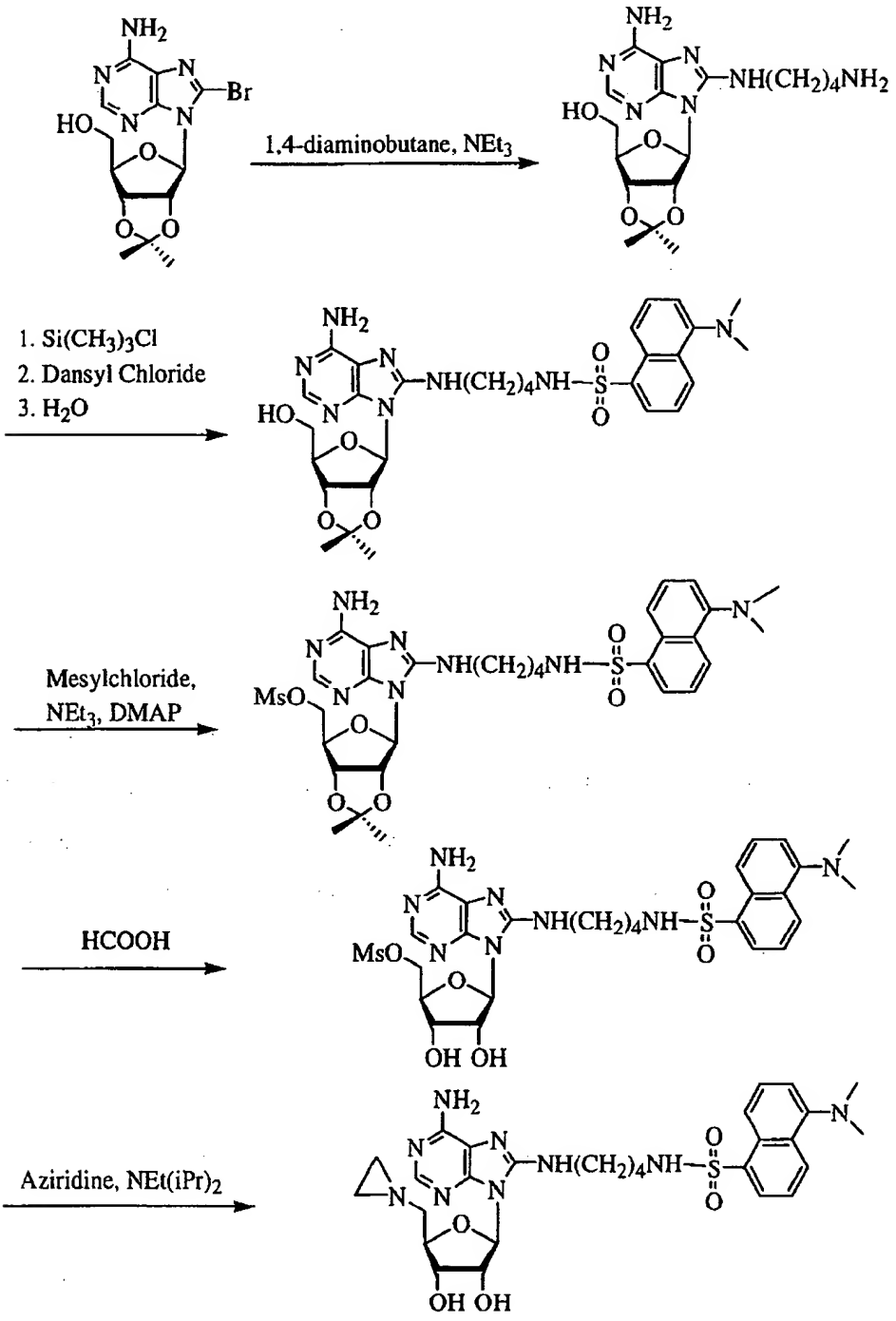
2. I am an expert in the field of synthetic chemistry and was an expert at the time of the invention. At the time of the invention I was employed as a group leader at Max-Planck-Gesellschaft zur Foerderung der Wissenschaften, assignee of the above-referenced patent application. Presently I am a Professor of Organic Chemistry at the RWTH Aachen (Rheinisch-Westfälische Technische Hochschule; Technical University of Aachen). My resume is attached as documentation of my credentials.

3. I declare that one skilled in the art at the time of the invention using the teaching of the specification, including the exemplary protocols as set forth in Examples 1 and 2, pages 19 to 30 of specification, and variations thereof, and other protocols known in the art at the time of the invention, could have successfully made and used the claimed compounds using only routine screening of alternatives.

As set forth in the specification, compounds of formula (I):

(I) 

can be prepared by the following exemplary Reaction Scheme, in which Y is N, R^1 is $-NH(CH_2)_4NHR^4$, R^4 is dansyl, and R^2 is hydrogen (see, *also,* Reaction Scheme 6 on page 14 of the specification):

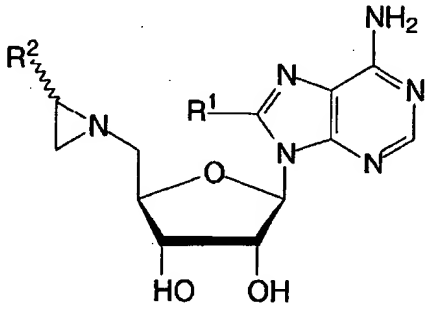


In particular, reaction of 8-bromo-2',3'-O-isopropylidene adenosine with 1,4-diaminobutane yields the protected adenosine derivative with an aminolinker at the 8 position (see, *e.g.*, Compound 1.1 of Example 2 in the specification). Transient protection of the 5'-hydroxy group with $Si(CH_3)_3Cl$, coupling of dansyl chloride with the primary amine of the aminolinker, and removal of the 5' hydroxyl protecting group leads

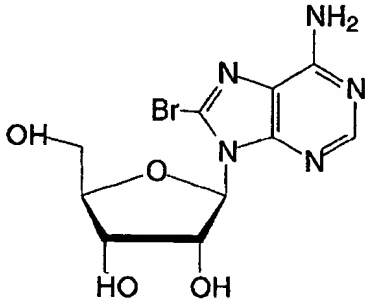
to the protected adenosine derivative with a fluorescent marker on the 8 position (see, *e.g.*, Compound 1.2 of Example 2 in the specification). This intermediate is then reacted with mesylchloride to yield the mesylate derivative (see, *e.g.*, Compound 1.3 of Example 2 in the specification). Removal of the isopropylidene protecting group under acidic conditions followed by reaction with aziridine affords a cofactor of formula (I).

In a similar manner, other compounds of formula (I) wherein Y is N, as represented below as formula (Ia), can be readily synthesized from a compound of formula (IIa) as set forth below according to the teaching of Reaction Scheme 6 and Example 2 of the specification as set forth above. The compound of formula (IIa), *i.e.*, 8-bromoadenosine, was commercially available from Aldrich Co. at the time the above-identified patent application was filed.

(Ia)



(IIa)

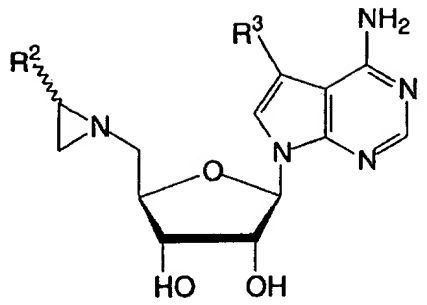


In particular, 8-bromoadenosine can be readily converted to 8-bromo-2',3'-O-isopropylidene adenosine under procedures well known to one of ordinary skill in the organic chemistry field. The bromo substituent at the 8-position can then be replaced by a diamine, such as $NH_2(CH_2)_nNH_2$ (where n is 1-3 or 4-250) or $NH_2(C_2H_5O)_nC_2H_5NH_2$ (where n is 1-250), under known amination conditions to form the appropriate intermediate corresponding to the intermediate prepared from 1,4-

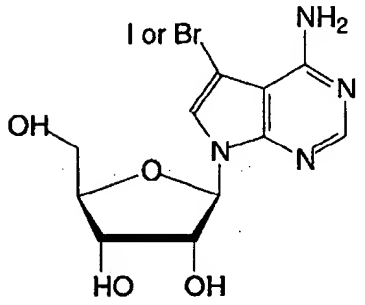
diaminobutane in Reaction Scheme 6. Such diamines would be considered by one skilled in the art to be homologues of $NH_2(CH_2)_4NH_2$ (1,4-diaminobutane) and, as such, would be expected to have comparable physiochemical properties to 1,4-diaminobutane in preparing the desired intermediates. The intermediates so prepared may then be treated with a compound of the formulae $XC(O)R^{4a}$ or $XS(O)_2R^{4a}$ (where X is bromo or chloro and R^{4a} is the rest of the R^4 group) under standard acylation or sulfonylation conditions to prepared compounds of the invention where R^4 is attached to an aminolinker. R^4 is defined by the specification as being common modifiers for biological molecules and that representative R^4 groups can be fluorophores, affinity tags, crosslinking agents, chromophores, proteins, peptides, amino acids, nucleotides, nucleosides, nucleic acids, carbohydrates, lipids, PEG, transfection reagents, beads and intercalating agents. Most, if not all of the compounds that could be used to afford the R^4 group to the claimed compounds are known to be reactive to free amine groups, and therefore can be easily reacted with the intermediate so formed in order to arrive at a compound of the invention. For example, the compound providing the R^4 group for the compounds of the invention in Reaction Scheme 6 is dansyl chloride, thereby forming a compound of the invention where R^4 is dansyl. Other compounds providing the R^4 group may be similarly reacted under conditions known to one skilled in the art with the intermediate to form compounds of the invention wherein R^1 is $-NH(CH_2)_nNHR^4$, or $-NH(C_2H_5O)_nC_2H_5NHR^4$, and R^4 is other than dansyl. The 5'-OH of the ribose moiety of the intermediate so formed can be activated with a good leaving group such as MsCl, as illustrated in Reaction Scheme 6, to form the corresponding –OMs group. Following deprotection of the hydroxyl groups of 2' and 3' position, 5'-OMs can then be replaced with aziridine to form a compound of formula (I).

Likewise, compounds of formula (I) when Y is $-CR^3$, as represented below as Formula (Ib), can be readily synthesized from a compound of formula (IIb) as set forth below. Compound of formula (IIb), *i.e.*, 7-iodo(or bromo)-7-deaza-adenosine, also called 5-iodotubercidin, is commercially available through Sigma-Aldrich.

(Ib)

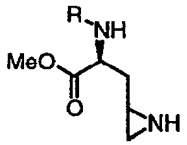


(IIb)



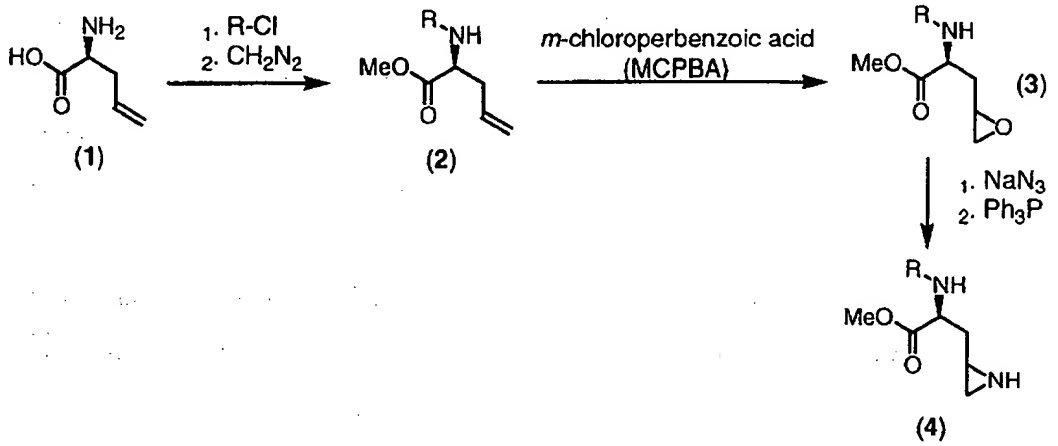
The halogen substitute (Br or I) at the 7 position of the compound of formula (IIb) is readily replaceable by diamines such as $NH_2(CH_2)_nNH_2$ and $NH_2(C_2H_5O)_nC_2H_5NH_2$ in a manner similar to that described above for compounds of formula (IIa), wherein Br at the 8 position is replaced with such a diamine. Consequently, compounds of formula (Ib) where R^3 is $-NH(CH_2)_nNHR^4$ or $-NH(C_2H_5O)_nC_2H_5NHR^4$, can be readily prepared by one skilled in the art according to the teaching of the specification and procedures and reagents known at the time the above-identified patent application was filed.

Furthermore, as one skilled in the art can readily appreciate, a compound of Formula (I) having a substituent on the aziridine ring can be synthesized according to Reaction Scheme 6 of the specification by replacing aziridine with an appropriately substituted aziridine. For example, when R^2 is $-CH_2CH(COOH)NH_2$, a suitably substituted aziridine precursor can be compound (4) as shown below:



(4)

Compound (4) can be readily prepared from a corresponding epoxide according a synthetic methodology known to one skilled in the art at the time of the invention. As exemplified by the following reaction scheme, L-allyl glycine (**1**, commercially available from Fluka) can be first protected to afford compound **2** in an art known manner. The double bond in **2** is then oxidized to provide an epoxide (**3**) in the presence of m-chloroperbenzoic acid, a well-known oxidizing agent. The expoxide (**3**) is then converted to the substituted aziridine (**4**) under a known reaction condition involving NaN_3 and Ph_3Ph (see, *e.g.*, Legters J. *et al.*, Tetrahedron Lett. **1989**, 30, 4881-4884).



Once compound (4) is attached to the adenosine component of Formula (I) according to the last step in Reaction Scheme 6, the aziridine component can be deprotected following art-known methods to provide a compound of Formula (I) wherein R^2 is $-CH_2CH(COOH)NH_2$.

4. I further declare that one skilled in the art could have used routine protocols known in the art at the time of the invention, including those described in the instant specification, to determine if any of the compounds of the invention acts as a co-factor for a S-Adenosyl-L-methionine (SAM) dependent methyltransferase.

In particular, Examples 1.3 and 2.2 of the specification provides detailed descriptions for preparing the enzymes M·TaqI and M·HhaI, and conducting enzymatic reactions, all of which are routine laboratory procedures known to one skilled in the art. Accordingly, a

screening protocol having general applicability based on these examples can be carried out in the following manner: The enzyme-catalyzed reaction can be carried out in a mixture (500 µl) of cofactor-free methyltransferase (5 nmol, 10 µM), a suitable substrate to the particular methyltransferase (5 nmol, 10 µM), a compound of formula (I) (10 nmol, 20 µM), Tris acetate (20 mM, pH 6.0), potassium acetate (50 mM), magnesium acetate (10 mM) and Triton X-100 (0,01 %) at 37°C. The progress of the reaction can be monitored by anion exchange chromatography (Poros 10 HQ, 10 µm, 4,6 x 10 mm, PerSeptive Biosystems, Germany). The product (which is the result of the enzyme-catalyzed transfer of the compound of formula (I) to the substrate) can then be eluted with aqueous potassium chloride (0.2 M for 5 min, followed by a linear gradient to 0.5 M in 5 min and to 1 M in 30 min) in Tris hydrochloride buffer (10 mM, pH 7.0).

Additionally, the specification, by way of detailed examples, provides three alternative means to analyze the product resulting from the transfer of a compound of formula (I) to a substrate in the presence of a suitable methyltransferase.

First, according to Example 1.3.1 on page 20 of the specification, the product can be analyzed directly by reversed-phase HPLC-coupled electrospray ionization mass spectrometry. More specifically, RP-HPLC/ESI-MS can be performed with an ion-trap mass spectrometer (LCQ, Finnigan MAT, Germany) equipped with a micro HPLC system (M480 and M300, Gynkotek, Germany). The product can be purified by anion exchange chromatography, followed by desalting by repeated addition of water and ultrafilteration (Microsep 3K, Pall Filtron, Northborough, MA, USA). The product solution can then be injected onto a suitable capillary column (for example, Hypersil-ODS, 3 mm, 150 x 0.3 mm, LC Packings, Amsterdam, Netherlands) and eluted with a linear gradient of acetonitrile (7-10% in 10 min, followed by 10-70% in 30 min, 150 µl/min) in triethylammonium acetate buffer (0.1 M, pH 7.0). The molecular weight obtained can be compared to the calculated molecular weight of the product.

Second, the product can be analyzed by electrospray ionization mass spectrometry using direct infusion according to Example 1.3.1 on page 21 of the specification. More specifically, a double focusing sector field mass spectrometer MAT 90 (Finnigan MAT, Germany) equipped with an ESI II electrospray ion source in the negative ion mode can be used. The desalted product in a aqueous solution and a liquid sheath flow (2-propanol) can then be delivered using a Harvard syringe pump (Harvard Apparatus, USA). The molecular weight of the product obtained from the electrospray mass spectra can then be compared to its calculated molecular weight.

Third, for a product wherein the substrate is an oligodeoxynucleotide (or oligonucleotide), the product can be analyzed by mass spectroscopy following an initial step of enzymatic fragmentation, according to Example 1.3.1 on page 21 and Example 2.2 on page 26. Specifically, a purified product of transferring a compound of formula (I) to an oligodeoxynucleotide (or oligonucleotide) can be dissolved in potassium phosphate buffer (10 mM, pH 7.0, 228 µl) containing magnesium chloride (10 mM), DNase I (2.7 U), phosphodiesterase from Crotalus durissus (0.041 U), phosphodiesterase from calf spleen (0.055 U) and alkaline phosphatase (13.7 U) and incubated at 37°C for 20 h. An aliquot (100 ml) can be injected onto a reversed-phase HPLC column (Hypersil-ODS, 5 mm, 120 Å, 250 x 4.6 mm, Bischoff, Leonberg, Germany), and the products can be eluted with a gradient of acetonitrile (0-10.5% in 30 min followed by 10.5-28% in 10 min and 28-70% in 15 min, 1 ml/min) in triethylammonium acetate buffer (0.1 M, pH 7.0). Beside the deoxynucleosides dC, dA, dG, T, and dA^{Me}, a new compound eluted can be found. The new compound can be isolated and detected by ESI-MS (LCQ connected to a nanoelectrospray ion source, Finnigan MAT, Germany). The observed mass is then compared with the calculated molecular mass of the product that is expected to be obtained by transferring a compound of formula (I) to the oligodeoxynucleotide substrate.

5. I declare that one skilled in the art could have used routine protocols known in the art at the time of the invention, including those described in the instant specification, to

determine if a putative methyltransferase could have complexed with an aziridine derivative of the present invention.

It was known in the art at the time of the invention that SAM-dependent methyltransferases for diverse substrates such as DNA, RNA, protein, peptide and even small molecules, had common catalytic domain(s) for binding the SAM cofactor. As demonstrated above, the instant invention is directed to the discovery that compounds of formula (I) behave in substantially the same manner as SAM in the presence of two particular DNA methyltransferase, M·TaqI and M·HhaI. See, for example, Reaction Scheme 7 of the specification. This result indicates to one skilled in the art that the compounds of the invention occupy the same catalytic domain of the DNA methyltransferases as SAM does, and would therefore function in the same manner for other SAM dependent methyltransferases due to the common catalytic domains of such enzymes.

5. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Respectfully submitted

Date: July 23, 2004

Elmar Weinhold

501346_1.DOC

CURRICULUM VITAE

Prof. Dr. Elmar Günther Weinhold

Address: Institut für Organische Chemie der RWTH Aachen
Professor-Pirlet Str. 1
D-52056 Aachen
Germany

Tel.: +49-241-80 94044
Fax: +49-241-80 92528
E.mail: elmar.weinhold@oc.rwth-aachen.de

Date of Birth: July 13, 1960

Place of Birth: Berlin, Germany

Positions/Education:

since 06.00 **Professor of Organic Chemistry** at the Institut für Organische Chemie der RWTH Aachen, Germany.

07.93 – 05.00 **Group leader** at the Department of Physical Biochemistry (*Prof. Roger Goody*), Max-Planck-Institut für molekulare Physiologie, Dortmund, Germany.
Habilitation in bioorganic chemistry at the Fachbereich Chemie, Universität Dortmund, Germany, with the title: Synthesis of modified duplex oligodeoxynucleotides and cofactor analogues for structure-function studies of DNA methyltransferases.

03.93 – 05.93 **Visiting scientist** at the New England Biolabs, Inc., Beverly, MA, USA.

03.91 – 02.93	**Postdoc** at the Department of Chemistry, Harvard University, Cambridge, MA, USA. Postdoctoral fellow of the Deutsche Forschungsgemeinschaft in the laboratory of *Prof. Jeremy R. Knowles* with the project title: Binding studies of synthestic sialic acid derivatives and influenza A hemagglutinin
07.86 – 02.91	**Ph.D.** at the Laboratorium für Organische Chemie der ETH-Zurich, Switzerland. Ph.D. thesis in the group of *Prof. Steven Benner* with the title: Protein engineering: A method for understanding the relationship between structure and activity of alcohol dehydrogenase from yeast.
02.86 – 06.86	**Graduate Student** at the Department of Chemistry, Harvard University, Cambridge, MA, USA.
03.80 – 01.86	**Diploma in Chemistry** (very good) at the Freie Universität Berlin, Germany. Diploma thesis with *Prof. Johann Mulzer* in the field of asymmetric synthesis.

Peer-reviewed publications

26. A. David, N. Bleimling, C. Beuck, J.-M. Lehn, E. Weinhold, M.-P. Teulade-Fichou, "DNA mismatch-specific base flipping by a bisacridine macrocycle", *ChemBioChem*, submitted.
25. C. Bolm, D. Müller, C. Dalhoff, C. P. R. Hackenberger, E. Weinhold, "The stability of pseudopeptides bearing sulfoximines as chiral backbone modifying element towards proteinase K", *Biorg. Med. Chem. Lett.*, in press.
24. C. Beuck, I. Singh, A. Bhattacharya, W. Hecker, V. S. Parmar, O. Seitz, E. Weinhold, "Aromatic DNA-base surrogates confer high-affinity binding to a native base flipping DNA methyltransferase", *Angew. Chem.*, in press.
23. C. Beuck, E. Weinhold, "Convenient synthesis of oligodeoxynucleotides containing 2'-deoxy-6-thioinosine", *Nucleosides, Nucleotides & Nucleic Acids* **2003**, *22*, 629–633.

22. G. Pljevaljcic, M. Pignot, E. Weinhold, "Design of a new fluorescent cofactor for DNA methyltransferases and sequence-specific labeling of DNA", *J. Am. Chem. Soc.* **2003**, *125*, 3486–3492.

21. J. Wölcke, E. Weinhold, "A DNA-binding peptide from a phage display library", *Nucleosides, Nucleotides & Nucleic Acids* **2001**, *20*, 1239–1241.

20. G. Vilkaitis, E. Merkiene, S. Serva, E. Weinhold, S. Klimasauskas, "The mechanism of DNA cytosine-5 methylation: Kinetic and mutational dissection of *Hha*I methyltransferase", *J. Biol. Chem.* **2001**, *276*, 20924–20934.

19. K. Goedecke, M. Pignot, R. S. Goody, A. J. Scheidig, E. Weinhold, "Structure of the *N*6-adenine DNA methyltransferase M·*Taq*I in complex with DNA and a cofactor analog", *Nature Struct. Biol.* **2001**, *8*, 121–125.

18. G. Vilkaitis, A. Dong, E. Weinhold, X. Cheng, S. Klimasauskas, "Functional roles of the conserved threonine 250 in the target recognition domain of *Hha*I DNA methyltransferase", *J. Biol. Chem.* **2000**, 275, 38722–38730.

17. A. N. Sharath, E. Weinhold, A. S. Bhagwat, "Reviving a dead enzyme: Cytosine deaminations promoted by an inactive DNA methyltransferase and an *S*-adenosylmethionine analogue", *Biochemistry* **2000**, *39*, 14611–14616.

16. M. Pignot, G. Pljevaljcic, E. Weinhold, "Efficient synthesis of *S*-adenosyl-L-homocysteine natural product analogues and their use to elucidate the structural determinant for cofactor binding of the DNA methyltransferase M·*Hha*I", *Eur. J. Org. Chem.* **2000**, *3*, 549–555.

15. B. Holz, E. Weinhold, "Higher binding affinity of duplex oligodeoxynucleotides containing 1,2-dideoxy-D-ribose to the *N*6-adenine DNA methyltransferase M·*Taq*I supports a base flipping mechanism", *Nucleosides & Nucleotides* **1999**, 18, 1355–1358.

14. B. Holz, N. Dank, J. E. Eickhoff, G. Lipps, G. Krauss, E. Weinhold, "Identification of the binding site for the extrahelical target base in *N*6-adenine DNA methyltransferases by photo-cross-linking with duplex oligodeoxyribonucleotides containing 5-iodouracil at the target position", *J. Biol. Chem.* **1999**, *274*, 15066–15072.

13. H. Pues, N. Bleimling, B. Holz, J. Wölcke, E. Weinhold, "Functional roles of the conserved aromatic amino acid residues at position 108 (Motif IV) and position 196 (Motif VIII) in base flipping and catalysis by the *N*6-adenine DNA methyltransferase from *Thermus aquaticus*", *Biochemistry* **1999**, *38*, 1426–1434.

12. M. Pignot, C. Siethoff, M. Linscheid, E. Weinhold, "Kupplung eines Nucleosids mit DNA durch eine Methyltransferase", *Angew. Chem.* **1998**, *110*, 3050–3053; "Coupling of a nucleoside with DNA by a methyltransferase", *Angew. Chem. Int. Ed.* **1998**, *37*, 2888–2891.

11. S. Serva, E. Weinhold, R. J. Roberts, S. Klimasauskas, "Chemical display of thymine residues flipped out by DNA methyltransferases", *Nucleic Acids Res.* **1998**, *26*, 3473–3479.

10. B. Holz, S. Klimasauskas, S. Serva, E. Weinhold, "2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases", *Nucleic Acids Res.* **1998**, *26*, 1076–1083.

9. H. Pues, B. Holz, E. Weinhold, "Construction of a deletion library using a mixture of
5'-truncated primers for inverse PCR (IPCR)", *Nucleic Acids Res.* **1997**, *25*, 1303–1304.

8. G. Schluckebier, M. Kozak, N. Bleimling, E. Weinhold, W. Saenger, "Differential binding of *S*-adenosylmethionine, *S*-adenosylhomocysteine and sinefungin to the adenine-specific DNA methyltransferase M·*Taq*I", *J. Mol. Biol.* **1997**, *265*, 56–67.

7. S. Serva, S. Klimasauskas, E. Weinhold, "Fluorescence studies of the DNA base flipping induced by a cytosine-5 methyltransferase", *Biologija* **1997**, 9–12.

6. S. A. Benner, T. M. Jermann, J. G. Opitz, S. A. Raillard, T. R. Zankel, K. Trautwein-Fritz, J. Stackhouse, M. I. Ciglic, M. Haugg, N. Trabesinger-Rüf, E. G. Weinhold, "Developing new synthetic catalysts. How nature does it", *Acta Chemica Scandinavica* **1996**, *50*, 243–248.

5. E. G. Weinhold, S. A. Benner, "Engineering yeast alcohol dehydrogenase. Replacing Trp54 by Leu broadens the substrate specificity", *Protein Eng.* **1995**, *8*, 457–461.

4. E. G. Weinhold, J. R. Knowles, "Design and evaluation of a tightly binding fluorescent ligand for Influenza A hemagglutinin", *J. Am. Chem. Soc.* **1992**, *114*, 9270–9275.

3. D. P. Baker, C. Kleanthous, J. N. Keen, E. Weinhold, C. A. Fewson, "Mechanistic and active-site studies on D(–)-mandelate dehydrogenase from *Rhodotorula graminis*", *Biochem. J.* **1992**, *281*, 211–218.

2. E. G. Weinhold, A. Glasfeld, A. D. Ellington, S. A. Benner, "Structural determinants of stereospecificity in yeast alcohol dehydrogenase", *Proc. Nat. Acad. Sci. USA* **1991**, *88*, 8420–8424.

1. J. Mulzer, U. Steffen, L. Zorn, C. Schneider, E. Weinhold, W. Münch, R. Rudert, P. Luger, H. Hartl, "Electrophilic additions to 3-*C*-[(methoxycarbonyl)methyl]-3-deoxy-D-ribofuranose enolates: A case of unusually efficient non-chelate-enforced chirality transfer", *J. Am. Chem. Soc.* **1988**, *110*, 4640–4646.

Book chapters

3. G. Pljevaljcic, F. Schmidt, A. Peschlow, E. Weinhold, "Sequence-specific DNA labeling using methyltransferases" in *Methods in Molecular Biology: Bioconjugation Protocols* (Ed.: C. M. Niemeyer), Humana Press, NY, in press.
2. B. Holz, E. Weinhold, "Probes for DNA base flipping by DNA methyltransferases" in *Bioorganic Chemistry: Highlights and New Aspects* (Eds.: U. Diederichsen, T. K. Lindhorst, B. Westermann, L. Wessjohann), Wiley-VCH, Weinheim, **1999**, pp. 337–345.
1. S. A. Benner, R. K. Allemann, A. D. Ellington, L. Ge, A. Glasfeld, G. F. Leanz, T. Krauch, L. J. MacPherson, S. Moroney, J. A. Piccirilli, E. Weinhold, "Natural selection, protein engineering, and the last riboorganism: Rational model building in biochemistry", *Cold Spring Harbor Symposia on Quantitative Biology* **1987**, *52*, 53–63.

Patent applications

2. M. Pignot, E. Weinhold, "New cofactors for methyltransferases", PCT patent application No. EP99/05405, date 28.07.1999.
1. M. Pignot, E. Weinhold, "New cofactors for methyltransferases", European priority application No. 98 11 4201.1, date 29.07.1998.

Published conference abstracts

14. K. Goedecke, M. Pignot, R. S. Goody, A. J. Scheidig, E. Weinhold, "Crystal structure of the *N*6-adenine DNA methyltransferase M·*Taq*I in complex with DNA and a cofactor analogue", *Biol. Chem.* **2001**, *382*, S170.
13. B. Lippert, M. Drumm, E. Weinhold, G. Kampf, M. B. L. Janik, R. K. O. Sigel, "Platinum complexes as probes for nucleic acid structures", *Book of Abstracts of the 6th International Symposium on Applied Bioinorganic Chemistry* **2001**, L9.
12. S. Serva, E. Weinhold, S. Klimasauskas, "Stopped-Flow fluorescence studies of DNA base flipping by *Hha*I methyltransferase", *Biochem. Soc. Trans.* **2000**, *28*, A468.

11. E. Merkiene, E. Weinhold, S. Klimasauskas, "Kinetics of cofactor binding and catalytic loop movements of *Hha*I methyltransferase", *Biochem. Soc. Trans.* **2000**, *28*, A464.

10. S. Serva, E. Weinhold, S. Klimasauskas, "Stopped flow fluorescence studies of DNA base flipping by the *Hha*I methyltransferase", *Biochimie* **1999**, *81*, 123.

9. S. Serva, E. Merkiene, G. Vilkaitis, E. Weinhold, S. Klimasauskas, "Energetic and kinetic aspects of DNA base-flipping by methyltransferases", *J. Biosci.* **1999**, *24*, 102.

8. E. Weinhold, M. Pignot, G. Pljevaljcic, "Sequence-specific Labeling of DNA Using New Cofactors for DNA Methyltransferases", *Biol. Chem.* **1999**, *380*, S207.

7. E. Weinhold, B. Holz, N. Bleimling, "Probes for DNA base flipping by DNA methyltransferases", *Biol. Chem.* **1998**, *379*, S142.

6. M. Pignot, E. Weinhold, C. Siethoff, M. Linscheid, "Micro HPLC/ESI MS^N of duplex oligonucleotides to study the methylation reaction of M·*Taq*I", *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics* **1998**, 267.

5. E. Weinhold, B. Holz, M. Pignot, H. Pues, J. Wölcke, "Probes for DNA base flipping by methyltransferases" in *2nd Workshop of Young European Bio-organic Chemists – WEB-98* (Eds.: L. A. Wessjohann, M. Kalesse), Prosciencia Verlagsbuchhandel Ulrich C. Philipp, Köln, **1998**, p. 18.

4. B. Holz, H. Pues, J. Wölcke, E. Weinhold, "Fluorescence studies on the base flipping mechanism of the DNA methyltransferase M·*Taq*I, *FASEB J.* **1997**, *11*, A1151.

3. J. Wölcke, M. Pignot, B. Holz, E. Weinhold, "Untersuchung des katalytischen Mechanismus der DNA-Methyltransferase aus *Thermus aquaticus* (M·*Taq*I)" in *6. Nachwuchswissenschaftler-Symposium Bioorganische Chemie* (Eds.: L. A. Wessjohann, T. K. Lindhorst, B. Westermann, U. Diederichsen), Prosciencia Verlagsbuchhandel Ulrich C. Philipp, Cologne, **1997**.

2. J. Wölcke, E. Weinhold, "Substrate specificity of the DNA methyltransferase from *Thermus aquaticus*: Influence of the 3'-neighbor base", *Biol. Chem. Hoppe Seyler* **1995**, *376*, S169.

1. E. G. Weinhold, A. Ellington, S. R. Presnell, G. M. McGeehan, S. A. Benner, "Evolution guidance: Engineering alcohol dehydrogenase and ribonuclease", *Protein Eng.* **1987**, *1*, 236–237.